Transforming Rural Economies Through Tertiary Education: Evidence from India *

Amparo Castelló-Climent† Abhiroop Mukhopadhyay‡ Ravinder‡

†*University of Valencia* ‡*Indian Statistical Institute, Delhi*

June 2026

Abstract

This paper analyzes whether tertiary-educated (skilled) individuals who remain in rural areas contribute to rural prosperity in India despite significant migration frictions that limit rural-to-urban mobility. Using village-level data, we estimate the causal impact of tertiary education on rural development using an instrumental variable. We find that a one percentage point increase in the share of tertiary-educated individuals raises village per capita consumption by about 7.2 percent, with robust results using night-light density as an alternative measure of prosperity. The findings remain consistent across a host of robustness checks and are not driven by alternative channels. Among the mechanisms at work, tertiary-educated individuals contribute to local prosperity by accessing productive private-sector employment in villages and nearby commutable towns - a conventional mechanism, while simultaneously improving agricultural productivity through greater access to information, agricultural technical advice, and crop diversification - a novel channel. The results highlight a distinct form of structural transformation in which migration frictions retain highly educated workers in rural areas, generating local productivity spillovers and raising rural incomes, suggesting that structural transformation occurs not only through migration to cities but can also occur within rural areas.

Keywords: Tertiary Education, Rural Development, Structural Transformation

JEL classification: I25, J24, O11, O13, O53

*We would like to thank the participants at the 2026 North American Summer Meeting of the Econometric Society, 2026 Asian Meeting of the Econometric Society, 40th Annual Meeting of the European Economic Association, 17th ACEGD, ISI Delhi, 27th ACCDE at Jadavpur University, 2023 CSH-AXA International Conference on Socio-Economic Inequality in India, 2022 AGW at Azim Premji University, and seminar participants at Warwick University, IIT Jodhpur, and Northeastern University in Boston. We would also like to express our appreciation for the financial support received from the Spanish Ministry of Economy and Competitiveness under the PID2022-138093NB-I00 MCIN project and the Planning and Policy Research Unit, ISI Delhi. This work was done during Ravinder's visit to UNU-WIDER, Helsinki, and he is thankful to Professor Kunal Sen for his guidance and mentorship.

1 Introduction

A central feature of economic development is the reallocation of labor out of agriculture and into higher-productivity manufacturing and service activities. This structural transformation process is closely associated with the accumulation of human capital: as education levels increase, workers acquire the skills that make them more productive outside of agriculture, and the resulting reallocation increases aggregate income (Caselli and Coleman II, 2001; Herrendorf et al., 2014; Porzio et al., 2022). The spatial dimension of this process typically involves a movement of educated workers from rural to urban areas, where high-skill employment is concentrated and where agglomeration economies amplify the returns to education (Gennaioli et al., 2013; Baum-Snow and Pavan, 2012). This account—higher education, sectoral reallocation, urbanization, growth—is well-supported by aggregate evidence across countries and regions. What it leaves open is a question that is quantitatively important for a large share of the world's population: what happens when migration frictions prevent the full reallocation of educated workers to cities? Do educated individuals who remain in rural areas contribute to local prosperity, and if so, through what channels? This paper presents causal evidence on these questions using data from approximately 427,000 villages across 27 states of India[1].

India offers an ideal setting for this investigation. Despite rapid growth in tertiary education, a large share of college-educated individuals continues to reside in rural areas – approximately one third of all tertiary-educated Indians lived in rural areas in 2011, and this phenomenon has become even more pronounced over time. As shown in Figure 1, the share of the rural population aged 25 to 34 with a university degree more than doubled between 2011 and 2023, rising from 6.9 to 16.6 percent.[2] This pattern reflects in part the substantial frictions that impede rural-to-urban migration in India[3]. Such frictions are however not limited to India. Many countries in Latin America, Asia and Africa show low intensity of such migration (Bell et al., 2015), leading to "among the largest sources of inefficiency in the global economy"(Munshi, 2020). In the case of India, with rising levels of education, the result is

[1] A detailed description of the village is given in Appendix subsection A.1.

[2] While we focus on the year 2011 in this cross sectional study, it is important to note that this phenomenon has become even more important over time.

[3] Munshi and Rosenzweig (2016) show that caste-based insurance networks create strong incentives to remain in the village even when urban wages are considerably higher, because migration severs access to the informal insurance these networks provide.

a large stock of tertiary-educated individuals living in rural villages—individuals whom the standard model of structural transformation would predict to be in cities—whose impact on local prosperity has not been systematically studied.

We address this question by estimating the causal impact of higher education on village prosperity, as measured by mean consumption. We source data from the Socio-Economic Caste Census of India (2011) and the Census of India (2011) for our main results.[4] The focus on the village is because it is the smallest spatial administrative unit in India, allowing us to incorporate possible externalities at the local level, in line with the evidence of college externalities in cities (Moretti, 2004).[5] The rationale to look at tertiary education is twofold: First, the skill misallocation inherent in tertiary educated population remaining in rural areas is apparent from a reading of the literature (Rossi, 2020). But an important first step in estimating this misallocation needs one to examine how this skilled population contributes to rural areas. Second, the literature on human capital on economic development of India has highlighted the role of tertiary education (Castelló-Climent et al., 2018; Castelló-Climent and Mukhopadhyay, 2013). It is therefore natural to ask whether such education contributes through those who live in the rural areas.

An important concern for identification is the endogeneity of human capital, as it evolves hand in hand with income levels and other determinants of development, such as institutions (Acemoglu et al., 2014). The level of education in a society often correlates with other positive institutional structures of the economy such as better judiciary, better property rights, etc. To address this concern, we use intra-state variation, since political and economic institutions, the judiciary and policies in India are determined by the central and state governments, which makes it unlikely that institutional differences drive our results. Further, we identify the causal effect using variation in the historical location of Catholic missions in India circa 1911, as an instrumental variable. An extensive reading of catholic history suggests that Catholicism in India was unregulated by the Vatican in the period from the 16th century to the 1930s. Hence, which region Catholic missions were located depended on a myriad of idiosyncratic strategies by individual missionaries. Some factors do predict the location of catholic missions, such as whether the district of their location was on the coast, whether the district had a railway

[4] We also use data from the National Sample Survey to look at some of our mechanisms.

[5] Moretti (2004) estimates a Mincerian equation augmented with average city education and finds evidence of spillovers from college education in U.S. cities.

line passing through, and whether the district was tribal. Most importantly, the locations were not determined by the income of the areas, as predicted by income tax collections. We posit that the location, conditional on controlling for predictors, is likely to be exogenous. However, catholic missionaries shifted the long-run supply of tertiary education by establishing educational institutions that became concentrated around the original mission locations (Castelló-Climent et al., 2018). The first stage of our IV regression confirms that villages closer to historical Catholic missions today have significantly higher tertiary completion rates. Further, we find that this stock of rural college graduates has a large and causal effect on village prosperity. A one percentage point increase in the share of the village population with tertiary education raises per capita consumption by 7.2 percent.

The validity of the causal attribution rests on the assertion that the exclusion condition is met; in other words, we claim that, conditional on a rich set of geographical and historical controls, the location of Catholic missions in 1911 affects current village prosperity only through its effect on tertiary education. We offer extensive evidence in support of this identifying assumption, showing that the instrument is uncorrelated with historical income, public good provision[6], current health infrastructure, rural migration rates, and a range of other potential confounders. The main result is robust to using night light density as an alternative measure of prosperity (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013, 2014; Alesina et al., 2016), to trimming and distributional transformations of the skewed education variable (Bellemare and Wichman, 2020), and to controlling for urbanization at the district and subdistrict level.

The mechanisms through which tertiary education raises village prosperity are an important contribution of the paper, and they reveal a form of structural transformation that the standard account does not predict. The expected channel—reallocation of educated workers from agriculture to the non-agriculture sector, in particular, higher-productivity non-farm private sector employment—is present in our data. Tertiary-educated rural adults are significantly more likely to have regular wage jobs in the private sector, whether within the village or in nearby commutable areas reached by daily commuting. Since private wages reflect marginal products more closely than public sector wages in developing countries (Pritchett, 2001), this

[6]The skills of decision makers at the local level could affect development Casey et al. (2023) and lead to higher provision of public goods.

finding implies that tertiary education increases both income and productivity in the non-farm sector. This channel is consistent with the standard structural transformation mechanism operating in an attenuated form: rather than migrating permanently to cities, educated workers commute to nearby towns, capturing some of the returns to non-farm employment while remaining in the village.

The more novel and surprising channel operates through agriculture. Despite the growth of the service sector, agriculture represented 54.6percent of rural employment in India in 2011. The standard model treats agriculture as the sector that educated workers leave, not as one that they improve from within. Yet we find that the presence of a tertiary-educated member in an agricultural household raises the gross value of agricultural output per unit of land, increases crop diversification,[7] and improves access to technical advice about agriculture. These findings are confirmed using satellite-based data on Net Primary Productivity[8], and are supported by heterogeneity analysis showing that the effect of tertiary education on village consumption is significant only in districts with good soil quality and adequate rainfall – precisely where the returns to better agricultural information and technology are highest. The agricultural channel is therefore not a residual or secondary finding. It represents a distinct mechanism through which human capital raises rural prosperity, one that operates within the agricultural sector rather than through reallocation away from it.

Interestingly, the effect of agricultural productivity does not operate only through tertiary-educated individuals who work directly in agriculture. In our sample, only 24 percent of college-educated rural adults are classified as agricultural workers. The effect also operates through educated household members who work outside agriculture, but serve as conduits for information about markets, input prices, and new technologies that the household applies to cultivation decisions. This intra-household informational spillover from tertiary-educated non-farmers to agricultural outcomes has not, to our knowledge, been previously documented. The existing literature on human capital and agricultural productivity—from the seminal contributions of Schultz (1975) and Foster and Rosenzweig (1995, 1996) on the Green Revolution, through the cross-country evidence of Gollin et al. (2021), to the reassessment of Reimers and Klasen (2013)—focuses on the education of the cultivator as the relevant margin. Our find-

[7]The merits of cross diversification have been shown in (Hazra, 2001; Joshi et al., 2004; Birthal et al., 2015)

[8]Using the Net Primary Product to measure agricultural productivity has been validated by Zaveri et al. (2020).

ings suggest that this framework misses an important channel through which human capital shapes agricultural productivity, and one that becomes more important as tertiary education penetrates rural areas and as returns to market information and technological knowledge in agriculture increase.

In summary, our findings describe a process of structural transformation that differs from the canonical account in two respects. First, it occurs within rural areas rather than through rural-to-urban migration. Educated workers who drive this process remain in the village due to migration frictions (Munshi and Rosenzweig, 2016). Second, it involves a within-agriculture productivity channel—driven by informational spillovers from educated household members—that the standard reallocation mechanism does not predict. The prosperity generated by the presence of educated workers is therefore a consequence of incomplete structural transformation rather than its completion, a distinction that has implications for how we think about the aggregate efficiency costs of migration frictions (Restuccia et al., 2008; Bryan et al., 2014) and about the design of policies aimed at raising rural incomes in developing countries where such frictions are large.

This paper is related to several strands of literature. First, it contributes to the literature on structural transformation (Herrendorf et al., 2014; Porzio et al., 2022; Caselli and Coleman II, 2001) by documenting a within-village channel through which human capital raises prosperity, one that the standard reallocation account, in which educated workers move permanently to cities, does not capture.

Second, it contributes to the literature on human capital and economic development by providing causal evidence at the village-level on returns to tertiary education in a large developing country. A substantial body of work shows that highly educated individuals are drivers of growth and regional prosperity (Gennaioli et al., 2013; Squicciarini and Voigtländer, 2015; Meisenzahl and Mokyr, 2011; Maloney and Caicedo, 2017; Castelló-Climent and Mukhopadhyay, 2013; Castelló-Climent et al., 2018). We extend this literature by identifying two channels through which tertiary education raises rural prosperity: a non-farm employment channel, whereby educated individuals access productive private-sector work in villages and nearby commutable towns, and an informational spillover channel within agriculture.

Third, it contributes to the literature on human capital and agricultural productivity (Schultz, 1964; Foster and Rosenzweig, 1995, 1996; Gollin et al., 2021; Reimers and Klasen,

2013), by documenting an intra-household channel, whereby tertiary-educated, not necessarily working as farmers, raise the productivity of cultivating household members through knowledge about markets, inputs, and technologies. This effect complements the cultivator-education margin studied in previous work.

Fourth, it contributes to the literature on migration frictions and misallocation (Munshi and Rosenzweig, 2016; Bryan et al., 2014; Restuccia et al., 2008; Munshi, 2020) by showing that the educated workers whom frictions keep in rural areas generate local productivity gains that may partially offset the aggregate efficiency cost of incomplete reallocation, a margin that standard misallocation accounting exercises do not capture.

The structure of the article is as follows. In section 2, we describe the data and set the empirical methodology. We discuss our identification strategy and present our main results in section 3. We describe several robustness checks in section 4. We analyze the mechanisms through which tertiary education influences rural prosperity in section 5. The conclusions reached are summarized in section 6.

2 Tertiary Education and Village Prosperity

2.1 Data and Econometric Model

For our empirical analysis, we use village-level data in India. According to the census, India had around 650 thousand villages in 2011. Our analysis amalgamates data on variables from different datasets that need to be merged, often based on villages' names, which are often recorded slightly differently across datasets. Since perfect matching is not possible for some villages, the final dataset covers about 450 thousand villages in 27 states.[9] The analysis focuses on the share of the population that has completed tertiary education, that is, those with a university degree and above. The definition of the variables and the source of the data used throughout the paper are documented in Table A5 in the Appendix.[10] We define *Share_ter*

[9]We match only those villages which are uniquely identified within the district or within the block. In the final analysis, we take only those villages that are perfectly matched.

[10]The source of the share of completed tertiary educated is the Socio-Economic and Caste Census (SECC). The SECC is different from the usual decennial census, since while the census only reports data on the number of literates at the village level, the SECC reports the composition in the population of different levels of education-primary, secondary, ... university.

as the share of the population who have completed tertiary education. [11] Figure 2 shows the spatial distribution of the share of tertiary education across subdistricts in rural India, with state boundaries overlaid. The lighter shades represent a lower share of tertiary education. The figure shows that southern and some northern Indian states have a higher share of people with tertiary education and that there is significant heterogeneity within states.

Estimating the effect of education on development at the village level allows us to incorporate externalities in a better way, despite some limitations such as the nonavailability of income per capita data at the village level. To overcome this limitation, we use different measures of development. We first use a consumption measure, estimated by Asher et al. (2021). This proxy is available as part of the SHRUG data library for village-level data. The measure is generated using small-area estimates following the methodology in Elbers et al. (2003).[12] As is standard in the literature, we use the log of mean consumption per capita as our dependent variable.[13]

The summary statistics for the main variables are reported in Table 1. The mean value of log consumption per capita is 9.66, and that of the share of tertiary education is 5.63 percent.[14] The unconditional correlation between tertiary education and consumption per capita is 0.3043. This correlation, however, could be driven by confounding factors correlated with both of them. To account for this, we include a broad set of current, geographical, and

[11] Ideally, we would want to consider in the numerator only those who are 25 and above and have completed tertiary education. However, the SECC reports data only on the total number of tertiary educated in the village. However, since tertiary education completes for most only around 22, the measurement error in this variable, though present, is unlikely to be large.

[12] This uses an imputation model based on the India Human Development Survey $2011-12$ and SECC 2011. The imputation model uses information on assets and earnings (see the Appendix page $6-7$ Asher and Novosad (2020)), at the household level, contained in both datasets. The SECC 2011 household data is only available to the SHRUG team, public data reports these measures for the village level, which we use in our paper.

[13] Since consumption is imputed, it has a distribution. For our analysis, we used the mean of the imputed value. However, 1000 draws of village-level per capita consumption are also reported in SHRUG which we use in the robustness check section 4.

[14] The share of tertiary educated may suffer from some measurement error as the maximum value is 100 %. The mean for the top 1% is 37.7 percent. We do not delete these high values as relatively less is known about the distribution of tertiary education in villages of India-hence there is no secondary data source to decide what the maximum values should be. In any case, our instrumental variable estimation strategy takes care of any measurement error issues. Our results are robust to trimming off the top 1%.

historical controls in our model and estimate the following specification:

$$ln_Cons_per_Capita_{vds} = \alpha_s + \theta Share_Ter_{vds} + \rho^{'} C_{vds} + \pi^{'} G_{vds} + \delta^{'} H_{ds} + \epsilon \qquad (1)$$

where v, d, and s denote village, district, and state, respectively. We eliminate the impact of omitted variables that vary at the state level by allowing a state-specific intercept term α_s; that is, dummy variables for each state. We estimate the model using robust standard errors and also cluster our standard errors at the sub-district level. This accounts for serial correlation between villages in the same sub-district, but more importantly, this is the level at which our instrument varies.[15]

We account for other observed differences by including the vector C of current variables. The controls include the village population and the population share of scheduled castes and tribes, which are historically disadvantaged groups, as described in the Indian constitution. These variables are taken from SECC 2011. The list of current controls is parsimonious by intent. We exclude most current variables because they are likely to be endogenous to factors influencing contemporary economic development. Instead, as described below, we take into account different initial conditions across districts through a broad set of historical controls.

We also take into account a vast empirical literature that has documented a strong correlation between geographical characteristics and current development (Sachs, 2003). We model log consumption per capita as a function of time-invariant characteristics with the vector of geographical controls, G. Specifically, we use the latitude and longitude of the centroid of the village and their squares, extracted from the village shape files available from Maps of India; the average length of rivers that pass through the district (in kms) and the average height (elevation) of the district (in kms), extracted using Arc GIS tools; annual rainfall and temperature in a village;[16] the area of the village; an indicator variable for districts with any coastal boundary, which is based on district maps from Census of India 2011; the proportion of soil that is sandy in a district, sourced from 'Soils of India' and is a data base of NBSS and LUP, Indian Council of Agricultural Research; and the distance to the nearest million plus city calculated from each village using Arc GIS.[17]

[15] In India, a subdistrict, also known as a taluka, tehsil, or mandal, is an administrative division within a district. It is the intermediary level of administration between the district and the village or town level. A subdistrict usually consists of several villages or towns grouped together for administrative purposes.

[16] This is available at http://catalogue.ceda.ac.uk/uuid/3f8944800cc48e1cbc29a5ee12d8542d

[17] According to the Census of India 2011, there are 53 million plus cities.

To capture differences in initial conditions across districts that may also impact current education and development, we include a rich set of historical variables, H. This set of controls, which are all at the district level, is sourced from Census of India in 1931 and includes the share of the district population that is urban, the share of population that is tribal, the share of the population that are Brahmans, an elite and educated social group; an indicator for whether the district was historically a part of Princely India and the presence of railways in 1909.[18,19] These variables account for the evolution of other contemporaneous variables.

2.2 Ordinary Least-Square Estimates

We start with a descriptive Ordinary Least-Square (OLS) regression of the log of village-level per capita consumption on the share of the adult population with tertiary education. The results are reported in Table 2, with each column adding more covariates. The first column shows a positive and statistically significant coefficient for the share of tertiary education. On average, villages with a higher proportion of college-educated individuals have higher prosperity, measured by per capita consumption.

This finding, however, could be driven by omitted variables related to both tertiary educated individuals and consumption per capita. The results in columns (2-5) show that this is not the case. The correlation between consumption per capita and tertiary education is robust to the inclusion of state-fixed effects and a wide range of current, geographical, and historical variables. It could also be possible that college-educated individuals are picking up the effect of lower levels of schooling. When we control for the share of the population with below graduate-level education our results remain unchanged column (6).[20]

[18] We take the data from the Census in 1931 because it is the first census year that provides reliable data for British India and Princely States at the district level. The crosswalk between districts in 1931 and 2001 was available in previous data based on constructed by Castelló-Climent et al. (2018) and we extend that to 2011 using Atlas of India: Census of India 2011.

[19] The data on railways is sourced from a map in the Indian Administrative Report on Railways, which is further processed by Arc GIS.

[20] We define the share of the population with below graduate-level (lower) education as follows:

$$\text{Share Lower Education (pop 05+)} = \frac{\text{Total population with some level of education but not graduate}}{\text{Total population aged (05+)}} \times 100$$

As mentioned earlier, we do not have an age-specific distribution of education levels. We estimate the share of

These findings lead to a strong correlation between the proportion of adults with college education and greater economic prosperity in Indian villages. Nevertheless, a challenge in such empirical analysis is the identification of the causal effect, since the level of education in a society often correlates with other positive institutional structures of the economy, such as a better judiciary system, better property rights, etc., which also impact the economy. It is therefore difficult to 'identify' the causal impact of education alone. This challenge could be addressed if we have an instrument for tertiary educated. The main idea is that the instrument accounts for the observed variation in the proportion of tertiary educated in the village, but has no direct effect on village-income levels. In the next section, we show that the historical location of Catholic missionaries serves our purposes.

3 Identification and Main Results

Our identification strategy uses the variation in the historic location of Catholic missions in India circa 1911 to analyze the impact of higher education on village prosperity. Catholic missions changed the supply and preference for higher education due to their emphasis on providing quality education. Although this change was achieved through an expansion of Catholic educational institutions after India's independence, they were located around the original location of Christian missions (Castelló-Climent et al., 2018). This was in contrast to Protestant missionaries whose main emphasis was on educating their congregation to read the Bible and, as a result, promoted literacy and lower levels of education (Mantovanelli, 2014).[21]

We source information on the location of Catholic missionaries from a map published in the first edition of Atlas Hierarchicus, which marks the name of every place in India where there was a Catholic mission or missionary in 1911. Catholicism was not a coordinated movement in India. While the early missionaries followed Portuguese conquest, missionaries also settled in the interior away from Portuguese strongholds. Our reading suggests that the individual preferences of missionaries played a role in addition to some predictable factors that we control.[22]

the population with a lower level of education based on the age group (05+), since a significant portion of the young population is currently enrolled in lower levels of education (primary, . . . , senior secondary).

[21] Protestant Missionary location can however not be used as an instrument as such location has now been found to correlate with a host of outcomes including women's empowerment. For more (see Calvi et al., 2022).

[22] Appendix subsection A.2 presents a brief historical overview of the expansion of Catholic missions in India and the institutional factors shaping their geographic spread.

Castelló-Climent et al. (2018) find that Catholic missionaries were more likely to be located in coastal districts, in districts with a railway presence, and in districts with a larger share of tribals. Reassuringly, the location of Catholic missionaries was uncorrelated with the share of Brahmans and income tax revenues per capita, a proxy for income.

We extend this idea to look at the location of the catholic mission relative to each village. To calculate the instrument, we use Arc GIS and information on the centroid of the village area and the exact location of the Catholic missions. To remove any possible bias specific to any particular village location, we use an aggregate measure as our instrument. The instrument is calculated as the mean distance of a village from the nearest catholic mission, averaged over all the villages in a subdistrict. We refer to this instrument as the mean distance from the Catholic mission and use its logarithmic value in our regressions.

We use the log of the mean distance from the Catholic mission in 1911 as an instrument for the share of tertiary educated in the village in 2011. We show that our instrument satisfies the two conditions to be a valid instrument. First, we provide evidence that the log of the mean distance to Catholic mission is highly correlated with the share of population that is tertiary education in the village. Since it is a historical instrument and might be less strong in some specifications, we report the Anderson-Rubin Test statistic to test for weak instruments, and we also provide Weak IV robust 95% confidence intervals to show that our results are bounded away from zero (Finlay and Magnusson, 2009).

Second, a critical requirement for identification is that the instrument has to influence the measure of village development only through the share of tertiary educated population in the village. Catholic missionaries could have located in richer and more educated places that could directly influence current development. As discussed above, while the district where the Catholic mission was located in 1911 was not entirely exogenous, some predictable variables explain their location.

In Table 3 we regress the distance of Catholic missions on a broad set of geographical and historical variables. Analogously to the results for the location of the district, our results at the subdistrict level indicate that the distance to Catholic missions is smaller in coastal areas and in districts with railway penetration in 1909.[23] However, understandably, the missionaries

[23]Railroad network during colonial India reduced the cost of trading and increased trade and agricultural income Donaldson (2018), as well as enlarged the size of the cities Fenske et al. (2023).

settled in cities and towns within districts, while they worked in rural areas around urban centers. Hence, the distance to cities and towns is positively correlated to the mean distance to catholic missions. When we control for the 1901 Census variables, we also find that such distances are smaller when the district is more urbanized. This is partly mechanical since the larger the urban sprawl or more the urban centers, the closer are the rural areas to it. Although this does reflect some positive selection, and we test its implication in many ways to rule out its independent impact, the fact that the mean distance is lower in tribal districts also reflects the opposite selection, giving further evidence to the idiosyncratic strategies of the catholic missionaries. Reassuringly, income tax per capita, a good proxy for income, does not correlate with our instrument, nor does the proportion of brahmins in a district. Even more significantly, institutions providing human capital schools and colleges have no correlation with our instrument.

Our identifying assumption then is that conditional on all these covariates, the mean distance to Catholic missions in 1911 is exogenous and can be used to examine the impact of tertiary education on village development. In the next section, we provide evidence showing that the closer the average village to a Catholic mission, the higher the share of the population with higher education in 2011. In section 4 we provide further evidence on the validity of our instrument.

Table 4 displays the results of IV estimates using the log of the mean distance to the Catholic missions as the source of exogenous variation. Panel (a) displays the first stage results. The coefficient of Catholic mission is negative and statistically significant at the 1 percent level. The coefficient remains significant when we control for state fixed effect, current, geographical, and historical variables in columns (2-5).[24] The results indicate that villages farther away from a Catholic mission in 1911 have a lower share of the population with tertiary education in 2011. The quantitative effect is large. A one-standard-deviation increase in the distance from a Catholic mission (46 km) reduces the proportion of tertiary educated in the village by 11 percentage points.

We use the exogenous source of variation in the share of tertiary educated to identify the

[24]The Kliebergen-Paap F-stat that tests for weak identification is high in columns (1-3). However, it falls as we add geographical and historical controls in columns (4-5). Hence, in panel (b) we report Anderson-Rubin Test statistics and weak IV robust 95% confidence interval, which provide evidence that the results do not suffer from a weak instrument problem.

causal effect of skilled educated on village prosperity. The results of the second stage are displayed in panel (b). The estimated coefficient of the share of tertiary education is positive and statistically significant in all specifications (column 1-5). However, as noted above, it could be argued that the share of tertiary education is picking up the impact of other levels of schooling. Similar to the OLS results, we address this issue by including the share of lower education. The IV coefficient of higher education is positive and significant, while the coefficient of the share of lower education is not statistically significant. Moreover, the coefficient of higher education is similar in magnitude to the estimate in column (5), strongly suggesting that omitting a lower level of education is not driving our results on tertiary education.[25]

Overall, the results indicate a strong and significant causal impact of tertiary education on village prosperity, as measured by log consumption per capita. The results hold with a large variety of controls and are not driven by lower levels of schooling. The quantitative effect is also large. A one percentage point increase in the share of tertiary educated in the village increases per capita consumption by 7.2 percent. In the next section we address various concerns regarding our estimates and provide a wide range of sensitivity checks to show the results are robust.

4 Robustness

4.1 Imputed consumption

The first concern is that our dependent variable is an imputed consumption measure rather than an actual consumption, and thus it is susceptible to estimation error. To address this concern, we plot a kernel density of the main IV specification (Silverman, 2018), utilizing 1000 draws of log consumption per capita provided by the SHRUG dataset. As depicted in

[25] In this specification, the share of the population with lower education is included as a control variable. Its purpose is to ensure that the estimated effect of tertiary educated on log consumption per capita reflects the returns to higher education and is not confounded by the distribution of education at lower levels. No causal interpretation is placed on its coefficient. As a robustness check, reported in Appendix Table A1, we instrument for both tertiary attainment and lower education simultaneously, using Catholic missionary distance and the share of primary schooling in 1991 as instruments. The effect of tertiary education on log consumption per capita remains positive and statistically significant under this specification. This suggests our main estimates are not sensitive to the inclusion of lower education as a control - whose coefficient, while strongly predictive in the first stage, remains statistically insignificant in the second stage.

Figure 3, the distribution of the coefficient remains positive and exhibits a range between 0.045 and 0.088. The mean and median of the distribution are similar: around 0.065. The IV coefficient of the main specification, including all controls in column (5) in Table 4, was 0.0723, which is close to the mean and median of the distribution. This evidence implies that our result is not specific to a chosen value of the imputed consumption.

4.2 Alternative measure of income

Mean consumption per capita is a model-generated prediction and may provide misleading predictions if the model is misspecified. We check the robustness of the results with nightlight density, an alternative measure of income that has been used in the literature (Henderson et al., 2012; Michalopoulos and Papaioannou, 2013; Alesina et al., 2016). The results are in Table 5. Panel (a) shows the results for the first-stage regression. The coefficient of log Catholic mission remains significant at the 1 percent level, with the F statistic ranging around 6 to 8. Additional tests, such as the Anderson-Rubin (AR) test and the weak IV robust 95 percent confidence interval indicate that our result is not subject to a weak instrument problem. In columns (1-3), we proxy rural income with the light density at night, measured by the Visible Infrared Imaging Radiometer Suite (VIIRS) annual data for 2015.[26]

Panel (b) shows the second stage results. The coefficient of the share of tertiary education is positive and statistically significant at 1 percent level with alternative measures of income. Column (1) shows the results with the full sample. However, in rural India, there is a problem with low light density at night, which results in many zero values and some outliers as well.[27] To address the issue of extreme values, we trim outliers in columns (2) and (3) by 1 percent and 2.5 percent (on both sides), respectively. The coefficients of the share of tertiary education remain positive and statistically significant. The economic effect when we use night light data

[26]The availability of VIIRS nightlight annual data at the pixel level for 2015 provides the closest approximation to 2011. In contrast to aggregation of monthly data, these data have undergone a screening process to exclude nights affected by stray light or clouds. Outliers due to ephemeral lights have also been removed, and background non-light values have been set to zero (Gibson et al., 2021; Gibson, 2021). It is important to note that the aggregation of monthly data may not accurately represent the data, as it does not take into account seasonal differences. The VIIRS nightlights are an improvement over DMSP-OLS (Wu and Wang, 2019), which is used in Chanda and Cook (2022), although they have used monthly data.

[27]Villages are considered to be at the lower end of the development spectrum, and as they are geographically small areas, the density of light at night is not well captured, leading to poor-quality data.

is very high and sensitive to trimming.

4.3 Skewed distribution

Another concern is related to the distribution of the main variables. The distribution of the share of tertiary education is skewed and has a considerable number of zero values. Likewise, the VIIRS nightlights also exhibit a skewed distribution, with many zeros and some outliers. This exacerbates the violation of the assumption of normality for the dependent variable. To address this issue, we transform these variables using an inverse hyperbolic sine (IHS) function. In cases with many zero values or outliers, the IHS transformation has been shown to be preferred over the natural logarithmic transformation (Bellemare and Wichman, 2020). They also provide a formula for computing the elasticity in this type of transformation.

The results, displayed in Table 6, show that the instrument is not weak, with an estimated coefficient that is statistically significant at the 5% level in all columns of panel (a). In panel (b), the coefficient of both the share of tertiary education and its IHS transformation are positive and statistically significant. However, the variables and their transformations cannot be interpreted in terms of the marginal effect. Instead, we have reported an elasticity measure that is comparable to various IHS transformations. The magnitude of the elasticity in the first four columns ranges from 0.363 to 0.437, while in columns (5-8) it ranges from 9.09 to 11.12, and are all bounded away from zero. Our results therefore pass this robustness check.

4.4 Geographical selection

Our results could be driven by the geographical selection of the placement of Catholic missionaries. To mitigate this issue, we restrict our sample to villages in the 75th percentile, 50th percentile, and 25th percentile, based on the distance to Catholic missions, similar to Calvi and Mantovanelli (2018). The results are presented in Table 7. The coefficient of the share of tertiary education remains positive and statistically significant in all cases, suggesting that the results are not driven by geographical selection.

4.5 Outliers and atypical observations

Our estimates could be influenced by a specific state. To address this concern, we plot a kernel density function of the IV coefficient of the share of tertiary education by dropping one state

at a time. The results of this analysis are presented in Figure 4 and in Figure A1 in the Appendix. As can be seen, the mode of the distribution is close to the coefficient of the main IV specification. Furthermore, the range of the distribution is between 0.0460 and 0.0952, which is above zero. This indicates that the estimates are not significantly affected by any particular state, which further supports the validity of our findings.

4.6 Trimming

Our results could also be driven by extreme values. In order to address this concern, we trim, in separate exercises, our dependent variable and main independent variable by one percent. This trimming process involves removing the top and bottom one percent of the values from these variables. The findings are displayed in Table 8. Both the first and second stages of our analysis remain statistically significant across all combinations of trimming the dependent and main independent variables. This indicates that our results are not driven by extreme values.

4.7 Exclusion restriction and other channels

In this analysis, we show that our results are not confounded by other factors related to missionary activity. The first set of potential confounders that we consider are health and medical interventions, since they often go hand in hand with those for education (Calvi and Mantovanelli, 2018). We look at the district level under five mortality rates and the supply of health centers at the village level, such as access to primary health centers and maternity centers. In addition, we look at some proxy for the quality of institutions in a district. They include the share of district land that is irrigated, violent crime rates, and a proxy for confidence in local political institutions (village councils). We also look at some amenities at the village level, such as the post office, a school index (a dummy variable based on the median value of principal component constructed using the indicator variable of primary, middle, secondary and higher secondary schools), and closed drainage.[28]

Our instrument would be invalid if Catholic missionaries affect village development through

[28]The source and description of all variables are given in Appendix Table A5. Confidence in local political institutions is a district-level average constructed from households surveyed in the India Human Development Survey. While this survey is not representative at the district level, it is a large survey covering 340 districts in India.

a mechanism different from that of tertiary education. To support the exclusion restriction, we first provide evidence that our IV is not correlated with these potential confounders. The results presented in Appendix Table A2 show that it is indeed the case. Second, we control for each of these potential confounders one by one in the main IV regression. The coefficient of the share of tertiary education remains positive and statistically significant in all specifications Table 9, which leads to the suggestion that our analysis is not confounded by these variables.

We noted earlier that missionaries settled in urban centers. Thus, a concern is that our results are driven by urbanization. To rule out this channel, we control for a wide variety of proxies related to urbanization. We control for the district urbanization rate, the subdistrict urbanization rate, the district-level urban wages for the tertiary educated relative to those for the non-tertiary educated, and the urban labor force participation for the tertiary educated at the district level. We present the results in Table 10. Our results hold, and the coefficient of tertiary education remains statistically significant after accounting for all of these factors related to urbanization. Although the villages closer to catholic missionaries were closer to urban centers, urbanization does not seem to drive the correlation between tertiary education and village prosperity.

5 Mechanisms

How does tertiary education affect village income? To answer this question, one would need to construct village-level accounts and decompose total village income into its constituent sources. However, such accounts do not exist for rural areas in India. Instead, we look at occupation profiles and, where possible, income in particular activities of interest, to give some evidence on the possible mechanisms. The canonical channel through which higher education raises income—and the one the structural transformation literature would predict—operates through the reallocation of educated workers from agriculture to higher-productivity non-farm employment (Herrendorf et al., 2014; Porzio et al., 2022). We document evidence for this channel: tertiary-educated rural adults are significantly more likely to hold regular wage jobs in the private sector, whether within the village or in nearby commutable areas. However, our more novel and perhaps more surprising finding is that tertiary education also raises agricultural productivity, a channel that has received little attention in the context of higher education. We discuss the agricultural channel first, precisely because it is the least anticipated

result, before turning to the non-farm employment channel.

5.1 The Agricultural Productivity Channel

The role of human capital in agricultural productivity has been debated extensively in the literature. The seminal contribution of Schultz (1964) argued that education was central to agricultural development by enabling farmers to deal with disequilibria and adopt new technologies. Foster and Rosenzweig (1995, 1996) showed, in the context of the Green Revolution in India, that the returns to primary education in agriculture were high precisely because new seed varieties required reading instructions, adjusting inputs, and learning from neighbours. Gollin et al. (2021) provides broader cross-country evidence that the productivity increases associated with the Green Revolution were concentrated in places with the human capital capacity to adopt and adapt new technologies. More recently, Reimers and Klasen (2013) revisit the micro-macro evidence on education and agricultural productivity and find that once education attainment—rather than enrollment or literacy—is used as the measure of human capital, both micro and macro estimates consistently show positive returns to education in agriculture. The literature on social learning in agriculture further shows that information about new technologies diffuses through social and geographic networks, and that farmers update their input decisions based on the experiences of those around them (Bandiera and Rasul, 2006; Conley and Udry, 2010; Porzio et al., 2022) emphasize that differences in human capital across workers are a central driver of the reallocation of labor out of agriculture during structural transformation, suggesting that human capital shapes not just productivity within agriculture but the composition of the agricultural workforce itself.

The existing literature has, however, focused predominantly on primary and secondary education as the relevant margin for agricultural productivity, particularly in the context of the Green Revolution and the adoption of high-yielding variety seeds. In that era, tertiary education completion rates in rural India were below two percent and largely confined to urban areas. The agricultural environment has since changed considerably. Mechanization has increased, markets have become more integrated, information about inputs, prices and new technologies has become more important, and the share of tertiary-educated individuals residing in rural areas has increased over time. Whether tertiary education matters for agricultural productivity in this more recent context, and through what channels, is a question the existing

literature has not addressed.

To explore this, we use data from the Situation Assessment Survey of Agricultural Households, conducted by the NSSO in 2013, which covers 29,652 households engaged in cultivation. Our independent variable is a dummy indicating the presence of at least one tertiary-educated member in the household. This allows us to capture effects from members who work directly in agriculture and from those who do not, but who may nonetheless provide the household with information about markets, prices, or new technologies. We use three proxies for agricultural productivity. Following Deolalikar (1981), we measure agricultural income as the gross value of agricultural produce per unit of cultivated land. We also examine crop diversification, measured as the share of crops cultivated by the household that are neither cereals nor pulses, given the evidence that diversification reduces downside risk in farming (Di Falco and Chavas, 2009; Di Falco et al., 2011). Finally, we use access to technical advice related to agriculture as a proxy for exposure to new technological knowledge. Since we can only identify the district where the farmer lives in our dataset, we construct the instrument so that it varies at the district level. We use the log of average distance to the Catholic missionary averaged over villages in the district and cluster standard errors at the district level.

The results, reported in Table 11, indicate that the presence of a tertiary-educated household member has a strong positive impact on all three outcomes. Households with a tertiary-educated member show higher agricultural gross value per unit of land (column 1), greater access to agricultural technical advice (column 2), and a higher degree of crop diversification (column 3). The Anderson-Rubin test statistics confirm that these results are bounded away from zero. Columns (4)-(6) show that the results are robust to including the share of lower schooling as an additional control, which confirms that the effects are not driven by lower levels of education within the household. The share of lower education is positively correlated to the outcomes. While these results are not causal estimates of lower education, they point out that lower education may also have a role to play in agricultural productivity, as shown in the literature on Green Revolution (Foster and Rosenzweig, 1996).[29] To further explore the sources of agricultural technical advice, Appendix Table A4 disaggregates advice by type. Two specific

[29]Similar to Appendix Table A1, we also instrument the share of lower education using the primary education share in 1991. The results, reported in Appendix Table A3, show that tertiary education remains statistically significant across all columns, while the share of lower education is neither positive nor significant in the second stage, even though the first stage is significant.

sources show positive and statistically significant effects: Radio/TV/Newspaper/Internet and the Veterinary Department. This is plausible, as tertiary-educated household members are more likely to access information on new agricultural technologies through media channels and formal institutional sources.

To verify these findings at a more aggregate level and address potential measurement error in household survey data on agricultural income, we complement our analysis using satellite-based data on Net Primary Productivity (NPP), a biophysical measure of the rate at which vegetation converts atmospheric carbon into biomass, which provides a direct proxy for agricultural productivity across diverse crop types (Zaveri et al., 2020). The data are aggregated to the subdistrict level, as are the rural tertiary completion rate; we use the instrument which has been used for the main results, as it varies at the subdistrict level. The results in Table 12 confirm the household-level findings: subdistricts with a higher share of tertiary-educated adults have significantly higher agricultural productivity, as measured by NPP. This result is robust across all specifications, from the most parsimonious in column (1) to the fully controlled specification in column (5).

The agricultural channel is further supported by the heterogeneity analysis in Table 13. The impact of tertiary education on per capita consumption is statistically significant only in villages where soil quality is good, and only where rainfall is above the median. This pattern is consistent with the view that the returns to human capital in agriculture are larger where agronomic conditions permit a wider range of input and technology choices. Where soil or rainfall constrain what can be grown, the marginal value of information about new technologies, diversification strategies, or market access is naturally lower.

These results are novel in an important respect. The existing literature on human capital and agricultural productivity—from the seminal work of Schultz (1964) and Foster and Rosenzweig (1996) on the Green Revolution, to the more recent reassessment by Reimers and Klasen (2013)—focuses on the education of the cultivator as the relevant margin. The implicit assumption is that human capital affects agricultural productivity by improving the decision-making capacity of the cultivator directly. Our findings suggest a different and hitherto unrecognised channel: even tertiary-educated household members who are not themselves engaged in agriculture raise the agricultural productivity of the household. As per the NSS 2009-10 Employment and Unemployment Round, only 29 percent of tertiary-educated individ-

uals are classified as working in market-oriented agricultural and fishery occupations. Yet the presence of a college-educated member in the household raises gross agricultural value per acre, increases crop diversification, and improves access to technical advice related to agriculture. This points to an informational externality operating within the household: tertiary-educated members serve as conduits for knowledge about markets, input prices, agronomic technologies, and institutional resources - knowledge that the household then applies to cultivation decisions. To our knowledge, this intra-household information spillover from tertiary-educated non-farmers to agricultural outcomes has not been previously documented in either the structural transformation literature or the agricultural productivity literature.

A natural question raised by these findings is why tertiary-educated individuals remain in rural villages at all, given the well-documented urban wage premium and the concentration of high-skill employment in cities. The answer lies in the significant frictions that impede rural-to-urban migration in India. Munshi and Rosenzweig (2016) show that caste-based insurance networks create strong incentives to remain in the village even when urban wages are considerably higher, because migration severs access to the informal insurance these networks provide. Bryan et al. (2014) shows more broadly in South Asia that migration is severely undersupplied relative to the social optimum, consistent with the presence of large non-pecuniary frictions beyond wage differentials alone. These frictions help explain the pattern we observe: tertiary-educated individuals remain in their villages, where they raise agricultural productivity through informational externalities within the household, and access non-farm employment through daily commuting to nearby towns rather than permanent migration.[30] Restuccia et al. (2008) show that barriers to labor mobility are a key driver of low agricultural productivity in developing countries. Our findings complement this by suggesting that when migration frictions keep educated workers in rural areas, their presence generates local informational spillovers that partially offset the efficiency cost of immobility - an implication that aggregate accounting exercises focused on sectoral misallocation do not capture.

[30]Note also that we have shown in column (6) of Appendix Table A2 that our instrument does not correlate with rural migration rates.

5.2 Non-Farm Employment Channel

The stylized model of human capital and sectoral growth posits that if the non-agricultural sector is more skill-intensive than the agricultural sector, a rise in education among the adult labour force will reallocate labour from farm to non-farm employment and raise income (Caselli and Coleman II, 2001; Herrendorf et al., 2014). The evidence on structural transformation typically associates this reallocation with migration from rural to urban areas (Lewis et al., 1954; Harris and Todaro, 1970; Gennaioli et al., 2013). Rossi (2020) surveys the broader macro-development literature and concludes that human capital accumulation is one of the most robust determinants of income differences across countries, with a substantial portion of that effect working through the reallocation of workers from low to high productivity sectors. Since we focus on tertiary-educated individuals who continue to reside in the village, migration and remittances are not the primary mechanisms at work in our setting, as the migration friction evidence above suggests.

To examine the non-farm channel, we first draw upon data from the village level primary census abstract data that provides village totals of those who are cultivators, agricultural laborers, household industry workers (small household manufacturing enterprises) and other workers that include government employees, teachers, factory workers, traders etc. We aggregate everyone except cultivators and agricultural laborers and call them non-farm workers. We calculate the percentage share of adults who are non-farm workers. Using this variable as our dependent variable in our main specification, we find that there is a positive and significant impact of higher share of tertiary educated adult population on the percentage of those in non-farm sector Table 14. This points to structural transformation, in terms of a move to the non-farm sector, albeit situated in the village.

We explore this further, drawing on employment data from the NSSO Employment-Unemployment Survey, 2009-10, covering 125,023 adults aged 25 and above in rural India. Occupation classes are defined differently as compared to the census. But it does allow us to classify workers into farmers, self-employed non-farm workers, agricultural laborers, daily wage non-farm workers, and regular wage workers in the non-farm private and public sectors. There is consensus in the literature that regular wage or salaried employment is on average the highest-paid category of occupation in India (Fulford, 2014). However, public sector wages in developing countries often reflect administrative pay scales rather than marginal productivity (Pritchett, 2001),

making private regular wage employment a better marker of productive non-farm employment than regular wage employment overall. This distinction also matters for interpretation: if the income gains from tertiary education reflected access to public sector employment rather than private sector productivity, it would be harder to attribute the village-level prosperity effect to a genuine increase in output.

We estimate the impact of being tertiary educated at the individual level, in addition to the district-level covariates used in our main village regressions. We use the individual and household-level controls - dummy for male, age, dummies for social group (SC/ST), an indicator variable that the household is Hindu, land ownership, household size, and proportion of adults in the labor force. As before, due to the lack of geographical identifiers below the district level, we use the instrument log of distance to the catholic missionary, averaged over villages in a district.[31] Panel (a) of Table 15 suggests that while our first stage is valid, our instrument may be weak. Hence, analogously to above, we test for weak instruments and report the 95% weak instrument robust confidence interval.

The results, displayed in panel (b) of Table 15, highlight the impact of tertiary education on obtaining a private regular job. A similar impact is not seen for public sector jobs and, therefore, not for regular jobs as a whole. This is not entirely surprising, as the share of regular jobs in the public sector is high and the public sector hires adults at various levels of education, such as matriculation, after higher secondary and after graduation. Interestingly, the fact that higher education, ceteris paribus, raises the probability of regular jobs in the private sector, but not in the public sector, suggests that the income increase is linked to an increase in productivity.

As the size of the village economy is limited, there is likely to be a relatively low supply of regular-wage jobs locally. In fact, commuting to work is common in rural areas and is more widespread than migration (Sharma and Chandrasekhar, 2014). We examine whether being tertiary educated increases the probability of individuals commuting daily to urban areas for work. The results in columns (4) and (5) show that although tertiary education does not increase the probability of commuting to urban areas for work, it increases the probability of commuting to undertake private work. Therefore, tertiary-educated people residing in rural areas are more likely to work in regular jobs in the private sector, whether the job is in the

[31] Standard errors are therefore clustered at the district level.

village or in the nearest town.[32]

Taken together, the results suggest a form of structural transformation that differs from the canonical account in which higher education leads to rural-to-urban migration and a shift toward service sector employment in cities.[33] In our setting, tertiary-educated individuals remain in their villages—held there partly by the migration frictions we have described—raise agricultural productivity through informational externalities within the household, and engage in productive private sector work in the local economy. It is this combination of channels, rather than remittance transfers from migrants, that accounts for the strong positive effect of tertiary education on village consumption per capita.

6 Conclusion and Discussion

The vast majority of the population in developing countries continues to reside in rural areas, making the question of what drives prosperity in these regions of utmost importance. In this paper, we examined the role of tertiary education in rural development in India and asked whether educated individuals who remain in rural areas contribute to local prosperity despite the migration frictions that limit the full reallocation of skilled labor to cities.

India offers a particularly relevant setting for this analysis because the expansion of tertiary education over the last few decades has increasingly reached rural areas, while a sizable share of college-educated individuals continues to reside outside cities. Although the literature on structural transformation has largely emphasized the movement of educated workers into urban non-farm employment, much less is known about the consequences of tertiary education for rural economies themselves. Our findings suggest that this lacuna is important.

The novelty of the paper is to provide causal evidence that tertiary education substantially raises rural prosperity. Using village-level data from India and exploiting historical variation in the location of Catholic missions as an instrument for tertiary education, we find that villages with larger shares of tertiary-educated individuals exhibit significantly higher levels of pros-

[32]Note that our robustness results have already shown, especially in the first stage, that impact of catholic missionaries on tertiary completion rates in the village is not driven by the local urbanization rates - the relation with Catholic missionaries and tertiary completion remains strong, even after accounting for urbanization rates.

[33]The migration modeled in Gennaioli et al. (2013) involves shifting to cities as the price of house rents in urban settlements is an important component of their model.

perity, measured by per capita consumption and night-light density. Our preferred estimates imply that a one percentage point increase in the share of tertiary-educated individuals raises village per capita consumption by approximately 7.2 percent. These findings remain robust to alternative specifications, transformations of skewed variables, trimming exercises, and extensive controls for geography, institutions, health interventions, infrastructure, lower levels of education, and urbanization.

The mechanisms through which tertiary education influences rural prosperity reveal a form of structural transformation that differs from the standard account in the literature. The conventional non-farm channel is present in our data. Tertiary-educated individuals are significantly more likely to participate in regular private-sector employment and to commute to nearby towns and urban areas for productive work. This suggests that tertiary education raises village incomes partly through greater participation in higher-productivity non-farm activities, even when migration to cities remains incomplete.

At the same time, we document a second and more novel mechanism operating within agriculture itself. Agricultural households with a tertiary-educated member exhibit higher agricultural productivity, greater crop diversification, and improved access to agricultural technical advice. These findings are confirmed using satellite-based measures of Net Primary Productivity and are strongest with better agro-climatic conditions, such as rainfall and soil quality, where the returns to information and technology adoption are likely to be higher. Importantly, many tertiary-educated individuals are not directly engaged in cultivation, suggesting that the agricultural effect operates through intra-household informational spillovers. Educated household members appear to facilitate access to information regarding technologies, markets, inputs, and cultivation practices, thereby improving agricultural decision-making and productivity.

Taken together, the findings point to a process of structural transformation that occurs not only through movement out of agriculture, but also through productivity improvements within agriculture generated by educated individuals who remain in rural areas. In this sense, the prosperity effects documented in the paper arise not from the completion of structural transformation through permanent urban migration, but from the presence of skilled individuals within rural economies themselves.

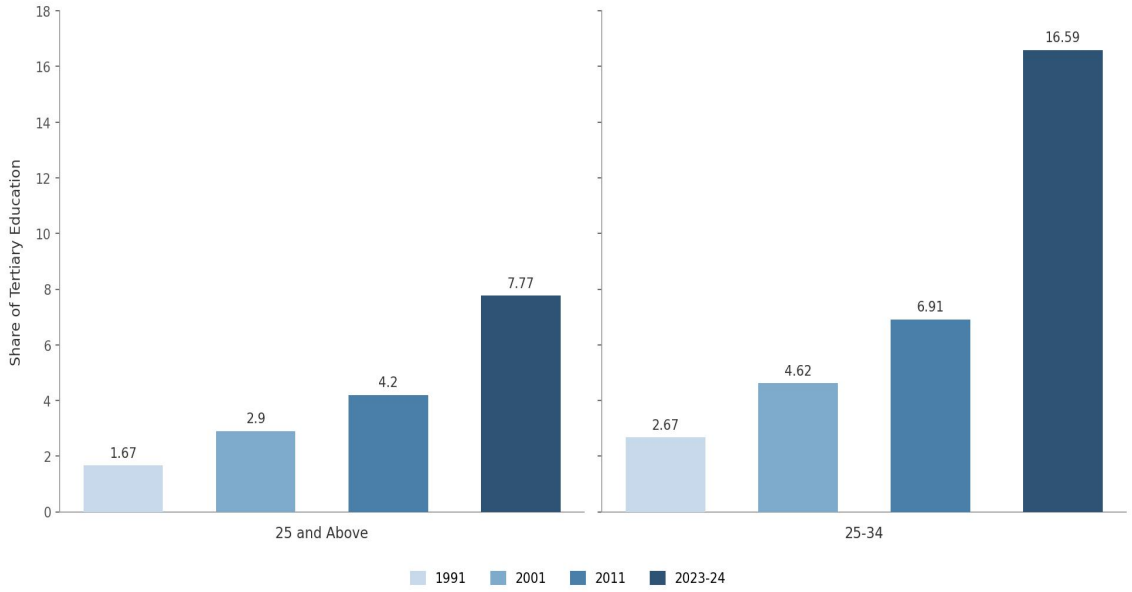
More broadly, the paper contributes to the literature on human capital, migration frictions,

economic development, and structural transformation. The results suggest that although migration barriers may generate aggregate inefficiencies by limiting the movement of educated workers to cities, the rural presence of tertiary-educated individuals also creates positive local spillovers that partially offset these costs. Human capital in rural areas raises productivity not only in the modern non-farm sector but also within traditional agriculture through the diffusion of knowledge and information.

Overall, the evidence presented in this paper suggests that tertiary education can play an important role in transforming rural economies in developing countries. In settings where migration frictions limit the movement of skilled workers to urban areas, educated individuals who remain in villages contribute to prosperity through improvements in both productive non-farm employment and agricultural productivity. The benefits of higher education therefore extend well beyond urban labor markets and constitute an important source of rural economic transformation. A weakness of our analysis, however, is that we have shown the occupation profile, but have not been able to decompose the village prosperity into its constituent parts fully.

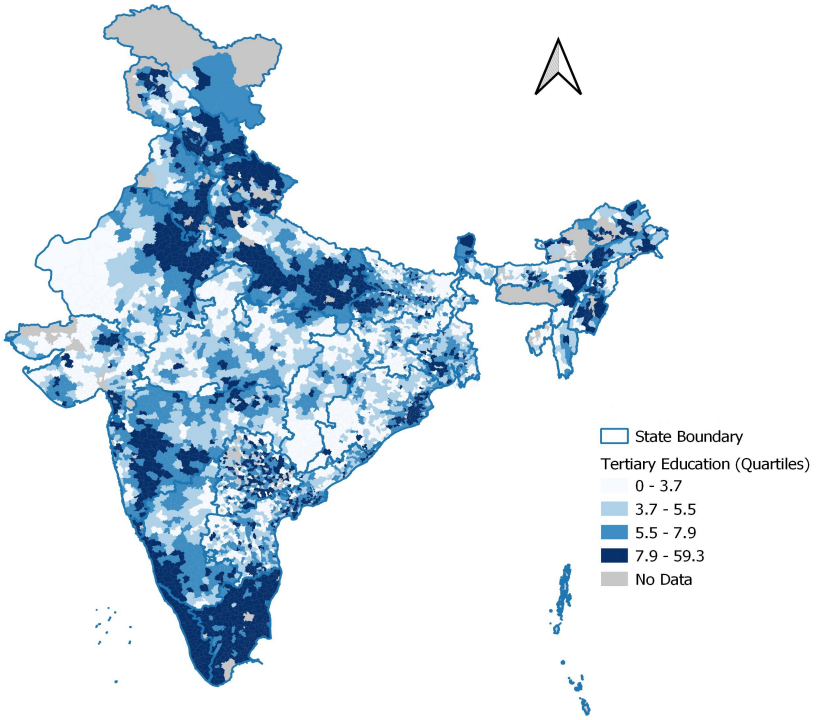
Tables & Figures

Figure 1: Share of Tertiary Education in Rural India from the Year 1991 to 2023-24



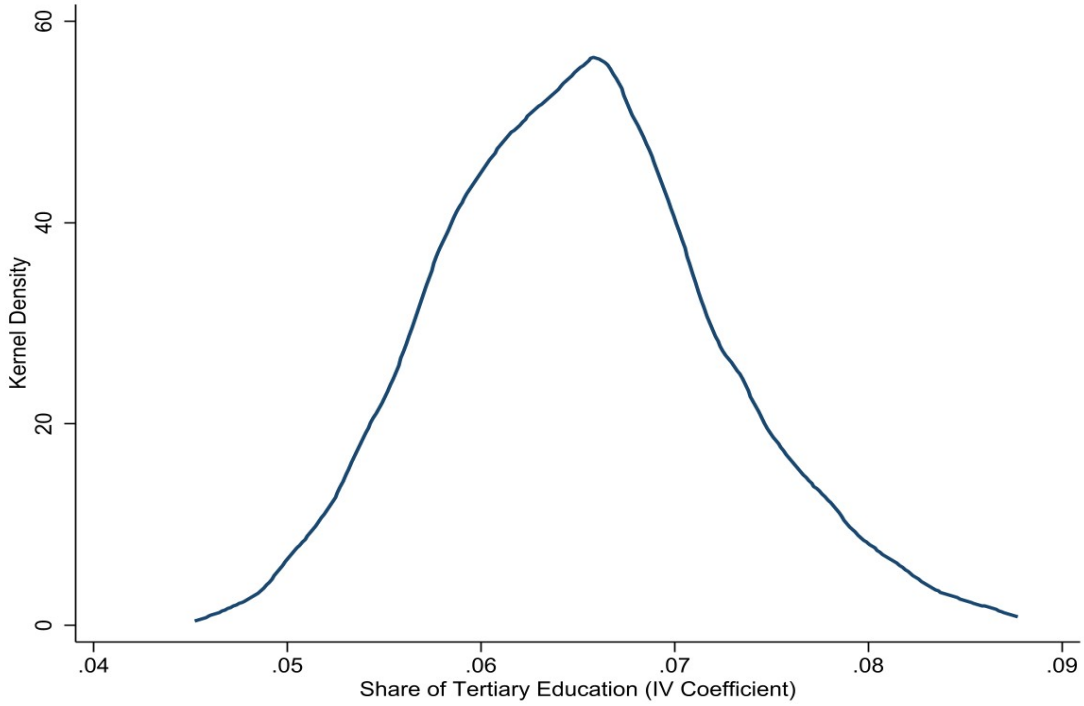
Notes: - The above graph shows the share of tertiary education in rural India over the last three decades for two age groups: individuals aged 25 or older and individuals aged 25 to 34. Data from the years 1991, 2001, and 2011 were obtained from the Population Census of India, while data from the year 2023-24 was obtained from the Periodic Labor Force Survey (PLFS), which is a nationally representative survey.

Figure 2: Share of Tertiary Education in Rural India Subdistrict Level



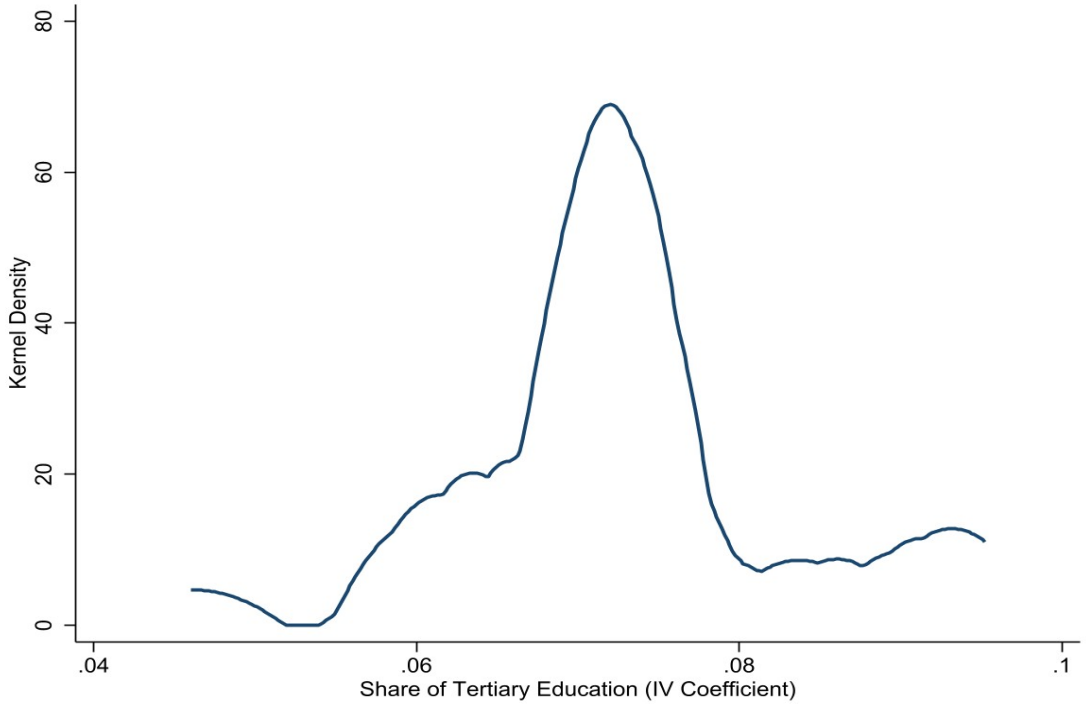
Notes: - Dark shading indicates a higher proportion of tertiary education in rural India, while light shading indicates a lower proportion.

Figure 3: Robustness, IV Estimates - 1000 Replications of Consumption per Capita (log)



Notes: - The kernel density plot displays the distribution of the IV coefficients for the share of tertiary education, using 1000 replications of bootstrapped log consumption per capita as the dependent variable.

Figure 4: Robustness, IV Estimates - Excluding One State



Notes: - The IV estimates use log consumption per capita as the dependent variable. The kernel density plot illustrates the distribution of coefficients for the share of tertiary education by iteratively excluding one state at a time.

Table 1: Summary Statistics

Variables	Observations	Mean	SD	Min	Max
Current					
Consumption per Capita (log)	428,960	9.662	0.290	8.951	10.60
VIIRS Nightlight *per Square km* Annual (log)	458,094	0.142	0.325	0	7.465
Share Tertiary & Above (pop 25+)	456,852	5.629	6.199	0	100
SC Proportion	458,472	0.178	0.209	0	1.000
ST Proportion	458,472	0.170	0.310	0	1.000
Total Population	458,472	1,431	1,868	0	51,108
Geography					
Latitude	457,967	23.64	4.732	8.094	35.34
Latitude Square	457,967	581.1	211.9	65.50	1,249
Longitude	457,967	80.76	5.152	68.52	97.07
Longitude Square	457,967	6,549	844.2	4,694	9,423
Coastal (dummy)	458,472	0.0838	0.277	0	1
Average River Length	457,634	12.68	3.585	2.932	30.34
Average Height District	458,472	370.2	502.2	3.967	4,942
Mean Rainfall (annual)	458,081	90.54	37.70	12.71	351.6
Soil Quality (sandy)	458,183	0.119	0.218	-1.19e-07	1
Nearest Distance from Big City	457,967	143.8	128.6	1.935	1,074
Mean Temperature (annual)	458,081	25.51	2.953	-4.275	29.74
Village Area Square km (log)	458,472	1.352	0.746	0	7.598
Historical					
Mean Subdistrict Distance Catholic Missionary	458,472	71.17	45.88	0	327.32
Fraction of Brahman (1931)	457,893	0.0588	0.0418	0	0.270
Fraction of Tribal (1931)	457,893	0.0448	0.107	0	0.829
Fraction of Urban (1931)	457,893	0.0955	0.0731	0	0.495
Princely State (dummy)	458,472	0.322	0.467	0	1
Rail (1909)	458,472	0.802	0.398	0	1
Historical (1901): Control in Table 3					
Fraction of Urban (1901)	343,228	0.0893	0.0871	0	0.468
Fraction of Brahman (1901)	343,228	0.0596	0.0433	0.00179	0.206
Fraction of Lower Castes (1901)	156,117	0.222	0.0964	0.00606	0.592
Fraction of Tribal (1901)	343,228	0.0545	0.113	0	0.957
Ethnic Fraction (1901)	156,117	0.725	0.149	0.317	0.898
Total Population (1901)	156,117	1.530e+06	719,100	111,437	2.913e+06
Colleges (1901)	156,117	2.462	3.794	0	17
School (1901)	156,117	954.9	904.2	0	4,558
Income Tax per Capita (1901)	156,117	0.0404	0.0287	0	0.364

Table 2: OLS Results, Log Consumption per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: - Consumption per Capita (log)						
Share Tertiary & Above (pop 25+)	0.0148***	0.0126***	0.0108***	0.0113***	0.0112***	0.0099***
	(0.0007)	(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Share Lower Education (pop 05+)						0.0034***
						(0.0001)
SC Proportion			-0.0720***	-0.0674***	-0.0706***	-0.0602***
			(0.0053)	(0.0050)	(0.0049)	(0.0046)
ST Proportion			-0.2758***	-0.2486***	-0.2467***	-0.1980***
			(0.0073)	(0.0075)	(0.0076)	(0.0070)
Population (log)			-0.0097***	-0.0154***	-0.0153***	-0.0184***
			(0.0012)	(0.0013)	(0.0013)	(0.0012)
Latitude				-0.1156***	-0.1076***	-0.1037***
				(0.0078)	(0.0077)	(0.0074)
Latitude Square				0.0028***	0.0026***	0.0026***
				(0.0002)	(0.0002)	(0.0002)
Longitude				-0.2117***	-0.1756***	-0.1941***
				(0.0288)	(0.0294)	(0.0262)
Longitude Square				0.0012***	0.0010***	0.0011***
				(0.0002)	(0.0002)	(0.0002)
Coastal (dummy)				-0.0097	-0.0185**	-0.0237***
				(0.0072)	(0.0076)	(0.0073)
Average River Length				-0.0004	-0.0005	-0.0014**
				(0.0006)	(0.0006)	(0.0006)
Average Height District				-0.0000	-0.0000	-0.0000
				(0.0000)	(0.0000)	(0.0000)
Mean Rainfall (annual)				0.0002**	0.0003***	0.0002***
				(0.0001)	(0.0001)	(0.0001)
Soil Quality (sandy)				-0.0035	-0.0016	-0.0273*
				(0.0158)	(0.0154)	(0.0142)
Nearest Distance from Big City				-0.0003***	-0.0002***	-0.0002***
				(0.0000)	(0.0000)	(0.0000)
Mean Temperature (annual)				0.0092***	0.0090***	0.0072***
				(0.0024)	(0.0024)	(0.0023)
Village Area Square km (log)				0.0160***	0.0156***	0.0207***
				(0.0025)	(0.0024)	(0.0022)
Fraction of Brahman (1931)					0.3426***	0.2412***
					(0.0919)	(0.0880)
Fraction of Tribal (1931)					0.0720***	0.0979***
					(0.0269)	(0.0254)
Fraction of Urban (1931)					0.2422***	0.1954***
					(0.0385)	(0.0371)
Princely State (dummy)					0.0213***	0.0322***
					(0.0053)	(0.0052)
Rail (1909)					0.0180***	0.0134**
					(0.0057)	(0.0054)
Observations	427,224	427,224	427,224	427,224	427,224	427,224
R-squared	0.092	0.375	0.433	0.456	0.460	0.491
State FE	NO	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: OLS Results, Determinants of Catholic Missionary Location

	(1)	(2)	(3)
Dependent Variable: - Mean Subdistrict Distance Catholic Missionary (log)			
Latitude	0.2771***	0.2363***	-0.1125
	(0.0854)	(0.0771)	(0.1712)
Latitude Square	-0.0062***	-0.0052***	0.0021
	(0.0020)	(0.0019)	(0.0045)
Longitude	1.5147***	1.2976***	0.8756
	(0.3321)	(0.3247)	(1.0147)
Longitude Square	-0.0096***	-0.0082***	-0.0055
	(0.0021)	(0.0020)	(0.0062)
Coastal (dummy)	-0.4484***	-0.4049***	-0.2956*
	(0.1098)	(0.1024)	(0.1760)
Average River Length	0.0048	0.0076	0.0037
	(0.0082)	(0.0082)	(0.0169)
Average Height District	0.0000	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0004)
Mean Rainfall (annual)	-0.0011	-0.0013	-0.0023
	(0.0012)	(0.0011)	(0.0021)
Soil Quality (sandy)	-0.0289	-0.0195	0.2213
	(0.1716)	(0.1680)	(0.4015)
Nearest Distance from City	0.0031***	0.0028***	0.0018**
	(0.0005)	(0.0005)	(0.0009)
Mean Temperature (annual)	0.0179	0.0219	0.0496
	(0.0198)	(0.0191)	(0.0447)
Village Area Square km (log)	0.0366	0.0342	-0.0372
	(0.0338)	(0.0339)	(0.0545)
Rail (1909)		-0.3363***	-0.4678***
		(0.0671)	(0.1213)
Fraction of Urban (1901)			-1.7704**
			(0.7258)
Fraction of Brahman (1901)			1.5436
			(2.2676)
Fraction of Lower Castes (1901)			0.8339
			(0.6566)
Fraction of Tribal (1901)			-1.0142*
			(0.5699)
Ethnic Fraction (1901)			0.0795
			(0.5884)
Total Population (1901)			-0.0000
			(0.0000)
Colleges (1901)			0.0200
			(0.0142)
School (1901)			-0.0000
			(0.0001)
Income Tax per Capita (1901)			-2.4251
			(1.4811)
Observations	5,309	5,309	2,375
R-squared	0.400	0.422	0.375
State FE	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the District Level, *** p<0.01, ** p<0.05, * p<0.1.

Table 4: IV Results, Log Consumption per Capita

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): First stage - Share Tertiary & Above (pop 25+)						
Mean Subdistrict Distance Catholic Missionary (log)	-1.0136***	-0.4531***	-0.4667***	-0.1956**	-0.2263**	-0.2082**
	(0.0864)	(0.0832)	(0.0819)	(0.0857)	(0.0880)	(0.0833)
Kleibergen-Paap rk Wald F Statistic	137.696	29.680	32.486	5.204	6.613	6.247
Panel (b): Second stage - Consumption per Capita (log)						
Share Tertiary & Above (pop 25+)	0.0765***	0.0516***	0.0576***	0.0940**	0.0723**	0.0717**
	(0.0074)	(0.0113)	(0.0116)	(0.0406)	(0.0287)	(0.0311)
Share Lower Education						0.0004
						(0.0016)
Anderson-Rubin (AR) Test Statistic Chi2(1)	175.56***	35.07***	63.88***	27.96***	21.01***	21.01***
Weak IV Robust 95% Confidence Interval	[0.0639, 0.0931]	[0.0341, 0.0825]	[0.0397, 0.0901]	[0.0506, ∞]	[0.0370, ∞]	[0.0360, ∞]
Observations	427,224	427,224	427,224	427,224	427,224	427,224
Controls						
State FE	NO	YES	YES	YES	YES	YES
Current Controls	NO	NO	YES	YES	YES	YES
Geographical Characteristics	NO	NO	NO	YES	YES	YES
Historical Variables	NO	NO	NO	NO	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 5: Robustness, IV Results - Alternative Measures of Income

	(1)	(2)	(3)
Panel (a): First stage - Share Tertiary & Above (pop 25+)			
Mean Subdistrict Distance Catholic Missionary (log)	-0.2577***	-0.2395***	-0.2237***
	(0.0863)	(0.0865)	(0.0868)
Kleibergen-Paap rk Wald F Statistic	8.915	7.676	6.641
Panel (b): Second stage			
	VIIRS Nightlight Annual (log)	VIIRS Nightlight Annual (log) Trim (1% both side)	VIIRS Nightlight Annual (log) Trim (2.5% both side)
Share Tertiary & Above (pop 25+)	0.1959***	0.1695***	0.1410***
	(0.0667)	(0.0614)	(0.0547)
Anderson-Rubin (AR) Test Statistic Chi2(1)	74.58***	75.07***	68.74***
Weak IV Robust 95% Confidence Interval	[0.1193, ∞]	[0.0989, ∞]	[0.0826, ∞]
Observations	454,915	450,376	443,569
Controls			
State FE	YES	YES	YES
Current Controls	YES	YES	YES
Geographical Characteristics	YES	YES	YES
Historical Variables	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 6: Robustness, IV Results - Inverse Hyperbolic Sine Transformation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): First stage								
	Share Tertiary & Above (pop 25+)	Share Tertiary & Above (pop 25+)	Share Tertiary & Above (pop 25+) (ihs)	Share Tertiary & Above (pop 25+) (ihs)	Share Tertiary & Above (pop 25+)	Share Tertiary & Above (pop 25+)	Share Tertiary & Above (pop 25+) (ihs)	Share Tertiary & Above (pop 25+) (ihs)
Mean Subdistrict Distance Catholic Missionary (log)	-0.2263**	-0.2263**	-0.0369**	-0.0369**	-0.2577***	-0.2577***	-0.0368**	-0.0368**
	(0.0880)	(0.0880)	(0.0154)	(0.0154)	(0.0863)	(0.0863)	(0.0150)	(0.0150)
Kleibergen-Paap rk Wald F Statistic	6.612	6.612	5.699	5.699	8.915	8.915	6.028	6.028
Panel (b): Second stage								
	Consumption per Capita	Consumption per Capita (ihs)	Consumption per Capita	Consumption per Capita (ihs)	VIIRS Nightlight Annual	VIIRS Nightlight Annual (ihs)	VIIRS Nightlight Annual	VIIRS Nightlight Annual (ihs)
Share Tertiary & Above (pop 25+)	1056.6137**	0.0723**			0.4290***	0.2500***		
	(431.9170)	(0.0287)			(0.1525)	(0.0852)		
Share Tertiary & Above (pop 25+) (ihs)			6486.5565**	0.4439**			3.0006**	1.7486**
			(2723.6265)	(0.1815)			(1.2753)	(0.7198)
Anderson-Rubin (AR) Test Statistic Chi2(1)	15.33***	21.03***	15.33***	21.03***	52.60***	74.33***	52.60***	74.33***
Weak IV Robust 95% Confidence Interval	[526.456, ∞]	[0.0371, ∞]	[3143.44, ∞]	[0.2354, ∞]	[0.2418, ∞]	[0.1522, ∞]	[1.6372, ∞]	[0.9790, ∞]
Elasticity	0.3625**	0.4070**	0.3893**	0.4371**	9.0902***	5.4810***	11.1203**	6.7051**
Observations	427,242	427,242	427,242	427,242	454,915	454,915	454,915	454,915
Controls								
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 7: Robustness, IV Results - Geographical Selection

	(1)	(2)	(3)	(4)
Panel (a): First stage - Share Tertiary & Above (pop 25+)				
Mean Subdistrict Distance Catholic Missionary (log)	-0.2263**	-0.3727***	-0.4992***	-0.6545***
	(0.0880)	(0.1141)	(0.1438)	(0.2198)
Kleibergen-Paap rk Wald F Statistic	6.612	10.669	12.050	8.866
Panel (b): Second stage - Consumption per Capita (log)				
	Full Sample	Distance Catholic Missionary $\leq$ 75 Percentile	Distance Catholic Missionary $\leq$ 50 Percentile	Distance Catholic Missionary $\leq$ 25 Percentile
Share Tertiary & Above (pop 25+)	0.0723**	0.0398***	0.0410***	0.0356**
	(0.0287)	(0.0152)	(0.0144)	(0.0154)
Anderson-Rubin (AR) Test Statistic Chi2(1)	21.03***	11.34***	13.42***	7.19***
Weak IV Robust 95% Confidence Interval	[0.0371, ∞]	[0.0175, ∞]	[0.0199, 0.0952]	[0.0118, ∞]
Observations	427,242	322,500	214,502	106,906
Controls				
State FE	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 8: Robustness, IV Results - Trimming

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel (a): First stage - Share Tertiary & Above (pop 25+)									
Mean Subdistrict Distance Catholic Missionary (log)	-0.2263**	-0.1694**	-0.2514***	-0.2271***	-0.2247**	-0.1809**	-0.2480***	-0.1682**	-0.2502***
	(0.0880)	(0.0761)	(0.0885)	(0.0873)	(0.0885)	(0.0758)	(0.0881)	(0.0764)	(0.0889)
Kleibergen-Paap rk Wald F Statistic	6.612	4.963	8.061	6.762	6.452	5.701	7.933	4.848	7.926
Panel (b): Second stage - Consumption per Capita (log)									
Share Tertiary & Above (pop 25+)	0.0723**	0.0953**	0.0594***	0.0736**	0.0738**	0.0919**	0.0622***	0.0972**	0.0600***
	(0.0287)	(0.0425)	(0.0228)	(0.0289)	(0.0295)	(0.0386)	(0.0238)	(0.0437)	(0.0229)
Trimming	No Trimming	X Trimmed by 1% from Top	X Trimmed by 1% from Bottom	Y Trimmed by 1% from Top	Y Trimmed by 1% from Bottom	X and Y Trimmed by 1% from Top	X Trimmed by 1% from Bottom and Y Trimmed by 1% from Top	X Trimmed by 1% from Top and Y Trimmed by 1% from Bottom	X and Y Trimmed by 1% from Bottom
Anderson-Rubin (AR) Test Statistic Chi2(1)	21.03***	20.63***	18.27***	23.30***	22.95***	23.22***	20.58***	22.49***	19.45***
Weak IV Robust 95% Confidence Interval	[0.0371, ∞]	[0.0464, ∞]	[0.0296, ∞]	[0.0381, ∞]	[0.0377, ∞]	[0.0477, ∞]	[0.0330, ∞]	[0.0505, ∞]	[0.0300, ∞]
Observations	427,242	423,280	380,862	423,036	422,956	419,444	377,266	419,014	378,490
Controls									
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: - X: Share Tertiary & Above (pop 25+) and Y: Consumption per Capita (log), Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 9: Robustness, IV Results - Alternative Channels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel (a): First stage - Share Tertiary & Above (pop 25+)											
Mean Subdistrict Distance Catholic Missionary (log)	-0.2096**	-0.2688***	-0.2445***	-0.2050*	-0.2474**	-0.2054**	-0.2365***	-0.2277***	-0.2199**	-0.2265***	-0.2262***
	(0.0853)	(0.0894)	(0.0894)	(0.1079)	(0.0997)	(0.0871)	(0.0869)	(0.0878)	(0.0869)	(0.0879)	(0.0878)
Kleibergen-Paap rk Wald F Statistic	6.038	9.031	7.470	3.610	6.158	5.564	7.411	6.723	6.407	6.650	6.645
Panel (b): Second stage - Consumption per Capita (log)											
Share Tertiary & Above (pop 25+)	0.0763**	0.0517***	0.0674***	0.0929**	0.0867**	0.0814**	0.0667***	0.0720**	0.0738**	0.0723**	0.0723**
	(0.0320)	(0.0190)	(0.0254)	(0.0466)	(0.0348)	(0.0346)	(0.0255)	(0.0284)	(0.0297)	(0.0286)	(0.0287)
Controls Included	Under 5 Mortality Rate	Share Irrigated Land	Violent Crimes per Capita	Confidence Panchayats	Agricultural Yield, 1956 (Log)	Rural Migration Rate	School Index	Primary Health Center	Maternity Center	Post Office	Drainage System
Anderson-Rubin (AR) Test Statistic Chi2(1)	20.58***	14.65***	21.06***	21.67***	34.71***	21.87***	19.90***	21.10***	20.54***	21.01***	21.09***
Weak IV Robust 95% Confidence Interval	[0.0395, ∞]	[0.0254, ∞]	[0.0362, ∞]	[0.0468, ∞]	[0.0468, ∞]	[0.0417, ∞]	[0.0354, ∞]	[0.0372, ∞]	[0.0373, ∞]	[0.0371, ∞]	[0.0371, ∞]
Observations	427,242	415,762	417,659	282,677	333,122	426,950	424,664	427,242	427,242	427,242	427,242
Controls											
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 10: Robustness, IV Estimates - Control for Factors Related to Urbanisation

	(1)	(2)	(3)	(4)
Panel (a): First stage - Share Tertiary & Above (pop 25+)				
Mean Subdistrict Distance Catholic Missionary (log)	-0.1933**	-0.1890**	-0.2781***	-0.2472***
	(0.0880)	(0.0902)	(0.0892)	(0.0898)
Kleibergen-Paap rk Wald F Statistic	4.823	4.389	9.712	7.574
Panel (b): Second stage - Consumption per Capita (log)				
Share Tertiary & Above (pop 25+)	0.0649**	0.0615**	0.0595***	0.0686***
	(0.0308)	(0.0307)	(0.0203)	(0.0257)
Urbanization Control	District Urbanization Rate	Subdistrict Urbanization Rate	Relative Urban Wage Tertiary vs. Non-Tertiary Education	Labour Force Participation Tertiary Educated (urban)
Anderson-Rubin (AR) Test Statistic Chi2(1)	12.12***	10.77***	20.33***	21.81***
Weak IV Robust 95% Confidence Interval	[0.0270, ∞]	[0.0239, ∞]	[0.0339, ∞]	[0.0370, ∞]
Observations	427,242	427,242	400,994	416,148
Controls				
State FE	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 11: Mechanism, IV Estimates - Agriculture Productivity Channel

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): First stage - Tertiary Educated Member in Household (dummy)						
Mean District Distance Catholic Missionary (log)	-0.0169***	-0.0173***	-0.0172***	-0.0189***	-0.0194***	-0.0193***
	(0.0065)	(0.0065)	(0.0065)	(0.0069)	(0.0069)	(0.0069)
Kleibergen-Paap rk Wald F Statistic	6.734	7.114	7.069	7.518	7.945	7.910
Panel (b): Second stage						
	Agriculture Output per Unit of Land (log)	Access to Technical Advice Related to Agriculture	Crop Diversification	Agriculture Output per Unit of Land (log)	Access to Technical Advice Related to Agriculture	Crop Diversification
Tertiary Educated Member in Household (dummy)	4.5246*	3.3604*	2.0398*	3.9384*	2.9375*	1.7955*
	(2.5875)	(1.7719)	(1.2290)	(2.2028)	(1.5320)	(1.0393)
Share Lower Schooling				0.0086**	0.0061**	0.0035*
				(0.0039)	(0.0027)	(0.0018)
Anderson-Rubin (AR) Test Statistic Chi2(1)	3.88**	5.82**	5.41**	3.76*	5.61**	5.32**
Weak IV Robust 95% Confidence Interval	[0.1191, ∞]	[0.7646, ∞]	[0.3367, ∞]	[0.0134, ∞]	[0.5718, ∞]	[0.2729, ∞]
Observations	29,457	29,596	29,652	29,457	29,596	29,652
Controls						
State FE	YES	YES	YES	YES	YES	YES
Individual & Household Controls	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors are in parentheses and clustered at the district level. *** p<0.01, ** p<0.05, * p<0.1. Household Controls: SC (dummy), ST (dummy), Hindu (dummy), Household Perform Livestock, Land Owned, and Land Owned Square. Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Distance from City, Mean Rainfall (annual), Soil Quality (sandy), Mean Temperature (annual), and District Population (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 12: Mechanism, IV Estimates - Net Primary Productivity (log)

	(1)	(2)	(3)	(4)	(5)
Panel (a): First stage - Share Tertiary & Above (pop 25+)					
Mean Subdistrict Distance Catholic Missionary (log)	-0.8274***	-0.4074***	-0.3898***	-0.1736***	-0.2271***
	(0.0646)	(0.0552)	(0.0546)	(0.0603)	(0.0620)
Cragg-Donald Wald F statistic	200.389	49.226	47.325	8.015	13.292
Kleibergen-Paap rk Wald F statistic	163.853	54.538	50.964	8.277	13.431
Panel (b): Second stage - Net Primary Productivity (log)					
Share Tertiary & Above (pop 25+)	0.0866***	0.2391***	0.2584***	0.4330***	0.3375***
	(0.0146)	(0.0363)	(0.0405)	(0.1540)	(0.0963)
Anderson-Rubin (AR) test statistic Chi2(1)	34.34***	143.51***	160.65***	8.277***	13.431***
Weak IV Robust 95% Confidence Interval	[0.0594, 0.1162]	[0.1831, 0.3297]	[0.1958, 0.3626]	[0.2562, ∞]	[0.2117, ∞]
Observations	5,300	5,300	5,300	5,300	5,300
Controls					
State FE	NO	YES	YES	YES	YES
Current Controls	NO	NO	YES	YES	YES
Geographical Characteristics	NO	NO	NO	YES	YES
Historical Variables	NO	NO	NO	NO	YES

Notes: - Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and District Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Distance from City, Mean Rainfall (annual), Soil Quality (sandy), and Mean Temperature (annual). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 13: Heterogeneity, IV Estimates

	(1)	(2)	(3)	(4)
Panel (a): First stage - Share Tertiary & Above (pop 25+)				
Mean Subdistrict Distance Catholic Missionary (log)	-0.3185***	-0.1003	-0.4945***	-0.0749
	(0.1039)	(0.1411)	(0.0882)	(0.1569)
Kleibergen-Paap rk Wald F statistic	9.406	0.505	31.410	0.228
Panel (b): Second stage - Consumption per Capita (log)				
Share Tertiary & Above (pop 25+)	0.0867***	-0.0049	0.0803***	-0.2311
	(0.0292)	(0.0570)	(0.0146)	(0.5096)
Weak IV Robust 95% Confidence Interval	[0.04845, ∞]	[-∞, ∞]	[0.0589, 0.1213]	[-∞, -0.0495] $\cup$ [$0.1119, \infty$]
Anderson-Rubin (AR) test statistic Chi2(1)	30.29***	0.01	81.29***	10.31***
Heterogeneity	Good Soil Quality	Bad Soil Quality	Good Rainfall	Bad Rainfall
Observations	216,303	210,939	212,397	214,845
Controls				
State FE	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 14: Mechanism, Tertiary Education and Non-Farm Employment

	(1)	(2)	(3)	(4)	(5)
Panel (a): First stage - Share Tertiary & Above (pop 25+)					
Mean Subdistrict Distance Catholic Missionary (log)	-1.0164***	-0.4648***	-0.4876***	-0.2201***	-0.2416***
	(0.0864)	(0.0831)	(0.0817)	(0.0840)	(0.0856)
Kleibergen-Paap rk Wald F Statistic	138.445	31.274	35.639	6.874	7.963
Panel (b): Second stage - Share Non-Farm Employment					
Share Tertiary & Above (pop 25+)	4.1452***	7.8217***	7.4324***	9.3041***	7.2240***
	(0.4133)	(1.4066)	(1.2782)	(3.5613)	(2.6774)
Anderson-Rubin (AR) Test Statistic Chi2(1)	131.04***	110.64***	115.41***	36.82***	26.14***
Weak IV Robust 95% Confidence Interval	[3.4415, 5.0453]	[5.7610, 11.8874]	[5.5598, 11.0257]	[5.2148, ∞]	[3.9376, ∞]
Observations	447,885	447,885	447,885	447,885	447,885
Controls					
State FE	NO	YES	YES	YES	YES
Current Controls	NO	NO	YES	YES	YES
Geographical Characteristics	NO	NO	NO	YES	YES
Historical Variables	NO	NO	NO	NO	YES

Notes: Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table 15: Mechanism, IV Estimates - Employment Channel

	(1)	(2)	(3)	(4)	(5)
Panel (a): First stage - Tertiary Educated (dummy)					
Mean District Distance Catholic Missionary (log)	-0.0077***	-0.0077***	-0.0077***	-0.0077***	-0.0077***
	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)
Kleibergen-Paap rk Wald F Statistic	6.717	6.717	6.717	6.717	6.717
Panel (b): Second stage					
	Regular Wage Earner * Private Sector	Regular wage Earner * Government Sector	Regular Wage Earner	Urban Work Location	Regular Wage Earner * Private Sector * Urban Location
Tertiary Educated (dummy)	0.6166**	-0.2431	0.3735	0.2836	0.2172*
	(0.2957)	(0.3100)	(0.3214)	(0.2742)	(0.1214)
Anderson-Rubin (AR) Test Statistic Chi2(1)	6.14**	0.90	1.04	1.07	4.54**
Weak IV Robust 95% Confidence Interval	[0.1600, ∞]	[-∞, 0.2111]	[-0.7336, 1.2516]	[-0.3569, ∞]	[0.0201, ∞]
Observations	125,023	125,023	125,023	125,023	125,023
Controls					
State FE	YES	YES	YES	YES	YES
Individual & Household Controls	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the District Level. *** p<0.01, ** p<0.05, * p<0.1. Individual & Household Controls: Gender, Age, Age Square, SC (dummy), ST (dummy), Hindu (dummy), Land Owned, Land owned Square, and Household Size. Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Distance from City, Mean Rainfall (annual), Soil Quality (sandy), Mean Temperature (annual), and District Population (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

References

Acemoglu, D., Gallego, F. A., and Robinson, J. A. (2014). Institutions, human capital, and development. *Annu. Rev. Econ.*, 6(1):875–912.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.

Asher, S., Lunt, T., Matsuura, R., and Novosad, P. (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in india using the shrug open data platform. *The World Bank Economic Review*, 35(4):845–871.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American economic review*, 110(3):797–823.

Bandiera, O. and Rasul, I. (2006). Social networks and technology adoption in northern mozambique. *The economic journal*, 116(514):869–902.

Banerjee, A. and Iyer, L. (2005). History, institutions, and economic performance: The legacy of colonial land tenure systems in india. *American economic review*, 95(4):1190–1213.

Baum-Snow, N. and Pavan, R. (2012). Understanding the city size wage gap. *The Review of economic studies*, 79(1):88–127.

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., and Kupiszewska, D. (2015). Internal migration and development: Comparing migration intensities around the world. *Population and Development Review*, 41(1):33–58.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Birthal, P. S., Roy, D., and Negi, D. S. (2015). Assessing the impact of crop diversification on farm poverty in india. *World Development*, 72:70–92.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748.

Calvi, R., Hoehn-Velasco, L., and Mantovanelli, F. G. (2022). The protestant legacy: missions, gender, and human capital in india. *Journal of Human Resources*, 57(6):1946–1980.

Calvi, R. and Mantovanelli, F. G. (2018). Long-term effects of access to health care: Medical missions in colonial india. *Journal of Development Economics*, 135:285–303.

Caselli, F. and Coleman II, W. J. (2001). The us structural transformation and regional convergence: A reinterpretation. *Journal of political Economy*, 109(3):584–616.

Casey, K., Glennerster, R., Miguel, E., and Voors, M. (2023). Skill versus voice in local development. *Review of Economics and Statistics*, 105(2):311–326.

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

Castelló-Climent, A. and Mukhopadhyay, A. (2013). Mass education or a minority well educated elite in the process of growth: The case of india. *Journal of Development Economics*, 105:303–320.

Chanda, A. and Cook, C. J. (2022). Was india's demonetization redistributive? insights from satellites and surveys. *Journal of Macroeconomics*, 73:103438.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, 100(1):35–69.

Deolalikar, A. B. (1981). The inverse relationship between productivity and farm size: a test using regional data from india. *American Journal of Agricultural Economics*, 63(2):275–279.

Di Falco, S. and Chavas, J.-P. (2009). On crop biodiversity, risk exposure, and food security in the highlands of ethiopia. *American Journal of Agricultural Economics*, 91(3):599–611.

Di Falco, S., Veronesi, M., and Yesuf, M. (2011). Does adaptation to climate change provide food security? a micro-perspective from ethiopia. *American journal of agricultural economics*, 93(3):829–846.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American economic review*, 108(4-5):899–934.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, 71(1):355–364.

Fenske, J., Kala, N., and Wei, J. (2023). Railways and cities in india. *Journal of Development Economics*, 161:103038.

Finlay, K. and Magnusson, L. M. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models. *The Stata Journal*, 9(3):398–421.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy*, 103(6):1176–1209.

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

Fulford, S. (2014). Returns to education in india. *World Development*, 59:434–450.

Gennaioli, N., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2013). Human capital and regional development. *The Quarterly journal of economics*, 128(1):105–164.

Gibson, J. (2021). Better night lights data, for longer. *Oxford Bulletin of Economics and Statistics*, 83(3):770–791.

Gibson, J., Olivia, S., Boe-Gibson, G., and Li, C. (2021). Which night lights data should we use in economics, and where? *Journal of Development Economics*, 149:102602.

Gollin, D., Hansen, C. W., and Wingender, A. M. (2021). Two blades of grass: The impact of the green revolution. *Journal of Political Economy*, 129(8):2344–2384.

Harris, J. R. and Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review*, 60(1):126–142.

Hazra, C. (2001). Crop diversification in india. *Crop diversification in the Asia-Pacific Region.(Minas K. Papademetriou and Frank J. DentEds.). Food and Agriculture Organization ofthe United Nations. Regional Office for Asiaand the Pacific, Bangkok, Thailand*, pages 32–50.

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

Herrendorf, B., Rogerson, R., and Valentinyi, A. (2014). Growth and structural transformation. *Handbook of economic growth*, 2:855–941.

Joshi, P. K., Gulati, A., Birthal, P. S., and Tewari, L. (2004). Agriculture diversification in south asia: patterns, determinants and policy implications. *Economic and political weekly*, pages 2457–2467.

Lewis, W. A. et al. (1954). Economic development with unlimited supplies of labour.

Maloney, W. F. and Caicedo, F. V. (2017). Engineering growth: Innovative capacity and development in the americas. 6339. *Publication Title: CESifo Working Paper Series. CESifo*.

Mantovanelli, F. (2014). The protestant legacy: Missions and literacy in india. *Available at SSRN 2413170*.

Meisenzahl, R. R. and Mokyr, J. (2011). The rate and direction of invention in the british industrial revolution: Incentives and institutions. In *The rate and direction of inventive activity revisited*, pages 443–479. University of Chicago Press.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Michalopoulos, S. and Papaioannou, E. (2014). National institutions and subnational development in africa. *The Quarterly journal of economics*, 129(1):151–213.

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

Munshi, K. (2020). Social networks and migration. *Annual Review of Economics*, 12(1):503–524.

Munshi, K. and Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(01):46–98.

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

Pritchett, L. (2001). Where has all the education gone? *The world bank economic review*, 15(3):367–391.

Reimers, M. and Klasen, S. (2013). Revisiting the role of education for agricultural productivity. *American Journal of Agricultural Economics*, 95(1):131–152.

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, 55(2):234–250.

Rossi, F. (2020). Human capital and macroeconomic development: A review of the evidence. *The World Bank Research Observer*, 35(2):227–262.

Sachs, J. (2003). Institutions don't rule: Direct effects of geography on economic development. *NBER Working Paper*, 9490.

Schultz, T. W. (1964). Transforming traditional agriculture.

Schultz, T. W. (1975). The value of the ability to deal with disequilibria. *Journal of economic literature*, 13(3):827–846.

Sharma, A. and Chandrasekhar, S. (2014). Growth of the urban shadow, spatial distribution of economic activities, and commuting by workers in rural and urban india. *World Development*, 61:154–166.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Squicciarini, M. P. and Voigtländer, N. (2015). Human capital and industrialization: Evidence from the age of enlightenment. *The Quarterly Journal of Economics*, 130(4):1825–1883.

Wu, K. and Wang, X. (2019). Aligning pixel values of dmsp and viirs nighttime light images to evaluate urban dynamics. *Remote Sensing*, 11(12):1463.

Zaveri, E., Russ, J., and Damania, R. (2020). Rainfall anomalies are a significant driver of cropland expansion. *Proceedings of the National Academy of Sciences*, 117(19):10225–10233.

Appendices

A Background

A.1 Village and Town Description

According to the 2011 Census of India, around 68% of Indians (833.1 million people) live in 640,867 different villages. The size of these villages varies considerably. Among all the villages in India, 236,004 have a population of fewer than 500, while 3,976 have a population of 10,000 and above. As per the 2011 Census, we don't have the definition of a village, but analogous to this, we have the definition of a town. The 2011 Census of India defines towns of two types: statutory towns and census towns. The statutory town is defined as all places with a municipality, corporation, cantonment board, or notified town area committee. Census towns are defined as places that satisfy three criteria: a minimum population of 5,000, at least 75% of the male working population engaged in non-agricultural pursuits, and a population density of at least 400/km^2 (1,000/mi^2). The towns in India usually have basic infrastructure like shops, electricity, bituminous roads, a post office, a bank, telephone facilities, high schools, and sometimes a few government offices. If any administrative unit does not satisfy the above criteria (census town or statutory town), it is considered a village.

A.2 Catholic Missionaries in India

We provide a short history of Catholic missionaries in India, based on Castelló-Climent et al. (2018). Church historical accounts put the entry of Catholic priests in India to 1498 with the arrival of Vasco de Gam and the Portuguese. Portuguese rulers enjoyed special ecclesiastical privileges, the Padroado Real, that gave them the right to fill clerical positions within overseas domains.

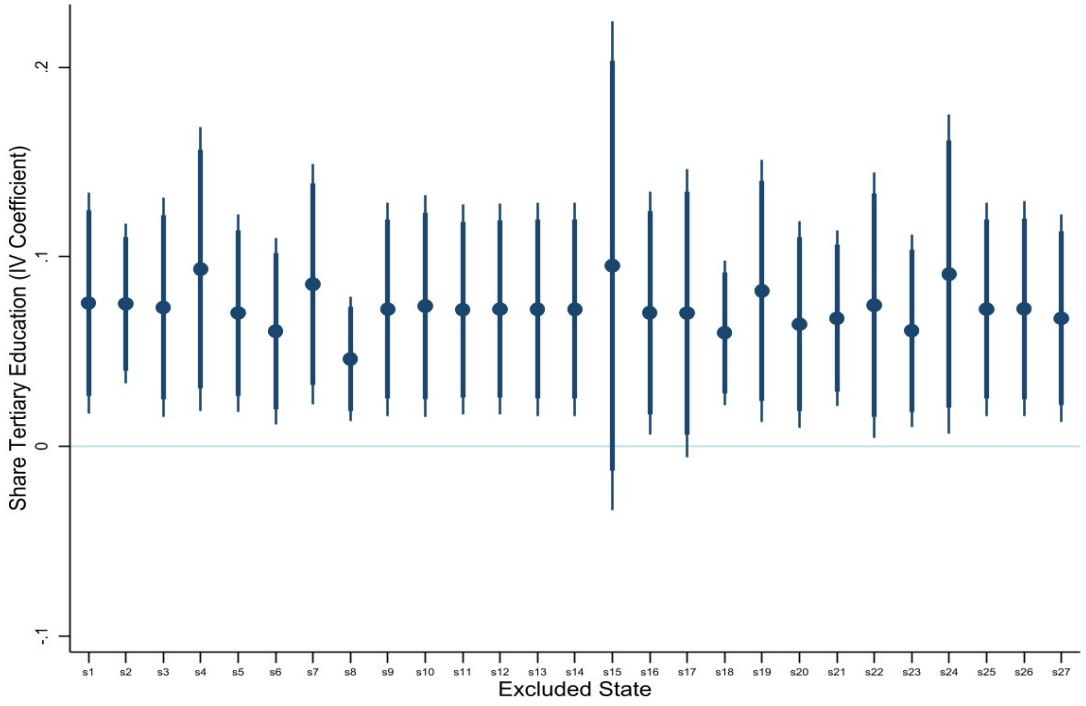
The first Catholic missions were set up in India in the 16th century. While the Franciscans and Dominicans dominated the scene early on, Jesuits played a much larger role, with the arrival of Francis Xavier in 1542. The initial missions were on the western coasts - Goa, Daman, Dui, Vasai (suburban area north of Mumbai) and Mumbai- which were Portuguese strongholds. Over-time, missionaries also settled in the interior parts of India. While penetration to the interior lands followed railways networks, it would seem that individual preferences

of missionaries played an important role in this expansion. For example, an enterprising Jesuit named Robert de Nobili moved to the cultural city of Madurai, pretended to be an upper caste Hindu and established the Madurai Mission to recruit high-caste Brahmans into the Catholic fold. There are accounts of Jesuit missionaries following Nobili's methods, as well as non-Jesuit missionaries working to convert lower castes and tribes. Akbar, the Mughal Emperor of India in this period, allowed Catholics to set up missions in Gujarat but the missionaries chose the location. Historical centers of trade and production were not always the natural choice.

As Portuguese rule declined over the 17th century, neither the native Indian rulers nor the English East India Company were sympathetic to the Catholic cause. The suppression of the Jesuits in the 18th century compounded the problem because they were the most active Catholic missionaries in the field.

Finally, ecclesiastical disputes between the Propaganda Fide, a society backed by Rome, and the Padroado Real backed by Portugal made it difficult for all Catholic missionaries. These differences were resolved by Pope Leo XIII in 1885 with a charter that established the Indian Catholic Church. Most jurisdictions were placed directly under Rome, or the Propaganda Fide barring two, the Archdiocese of Goa and the Diocese of Mylapore, near modern-day Chennai.

Figure A1: Robustness, IV Estimates - Excluding One State



Notes: The bold line represents the 90% confidence interval, while the thinner line represents the 95% confidence interval. The IV estimates use log per capita consumption as the dependent variable. The states that are excluded in the analysis are Jammu and Kashmir, Himachal Pradesh, Punjab, Uttarakhand, Haryana, Rajasthan, Uttar Pradesh, Bihar, Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Assam, West Bengal, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Gujarat, Maharashtra, Andhra Pradesh, Karnataka, Goa, Kerala, and Tamil Nadu, in that order.

Table A1: Robustness, Instrument Lower Level of Education

	(1)
Panel (a): First stage - Share Tertiary & Above (pop 25+)	
Mean Subdistrict Distance Catholic Missionary (log)	-0.2560***
	(0.0874)
Share Primary Schooling 1991	18.1432***
	(1.2079)
Panel (b): First stage - Share Lower Education	
Mean Subdistrict Distance Catholic Missionary (log)	-0.5744**
	(0.2462)
Share Primary Schooling 1991	128.914***
	(4.5846)
Kleibergen-Paap rk Wald F statistic	2.278
Panel (c): Second stage - Consumption per Capita (log)	
Share Tertiary & Above (pop 25+)	0.0837**
	(0.0419)
Share Lower Education	-0.0048
	(0.0060)
Anderson-Rubin (AR) Test Statistic Chi2(2)	217.11***
Observations	422,510
Controls	
State FE	YES
Current Controls	YES
Geographical Characteristics	YES
Historical Variables	YES

Notes: - Robust standard errors are in parentheses and clustered at the Subdistrict Level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual) and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table A2: OLS Results, Catholic Missionary on Alternative Channels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Under 5 Mortality Rate	Share Irrigated Land	Violent Crimes per Capita	Confidence Panchayats	Agricultural Yield, 1956 (Log)	Rural Migration Rate	School Index	Primary Health Center	Maternity Center	Post Office	Drainage System
Mean Subdistrict Distance Catholic Missionary (log)	-0.0862	-0.6599	-0.0000	-0.0028	-0.0216	-0.0005	-0.0033	-0.0001	-0.0060	-0.0009	-0.0012
	(1.0164)	(0.7403)	(0.0000)	(0.0092)	(0.0194)	(0.0006)	(0.0033)	(0.0008)	(0.0042)	(0.0013)	(0.0028)
Observations	456,529	444,855	444,268	302,716	353,437	456,230	449,872	456,529	456,529	456,529	456,529
R-squared	0.782	0.570	0.404	0.486	0.643	0.242	0.273	0.070	0.089	0.098	0.241
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Current Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: - Robust standard errors are reported in parentheses. In columns (1) to (6), the dependent variable varies at the district level, so we cluster standard errors at the district level. In columns (7) to (11), we cluster standard errors at the subdistrict level. *** p<0.01, ** p<0.05, * p<0.1. Current Controls: SC Proportion, ST Proportion, and Population (log). Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Mean Rainfall (annual), Soil Quality (sandy), Nearest Distance from City, Mean Temperature (annual), and Village Area Square km (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table A3: Mechanism, Agricultural Productivity Channel - IV Estimates Lower Education

	(1)	(2)	(3)
Panel (a): First stage - Tertiary Educated Member in Household (dummy)			
Mean District Distance Catholic Missionary (log)	-0.0198***	-0.0203***	-0.0202***
	(0.0065)	(0.0065)	(0.0065)
Share Primary Schooling 1991	0.2352***	0.2338***	0.2354***
	(0.0836)	(0.0834)	(0.0834)
Panel (b): First stage - Share Lower Schooling			
Mean District Distance Catholic Missionary (log)	-1.2899	-1.3257	-1.3319
	(0.9221)	(0.9187)	(0.9205)
Share Primary Schooling 1991	97.4257***	97.6056***	97.6135***
	(11.1741)	(11.0379)	(11.1171)
Kleibergen-Paap rk Wald F Statistic	2.819	2.969	2.926
Panel (c): Second stage			
	Agriculture Output per Unit of Land (log)	Access to Technical Advice Related to Agriculture	Crop Diversification
Tertiary Educated Member in Household (dummy)	4.4345*	4.2265*	1.9008^{+}
	(2.6438)	(2.3106)	(1.2590)
Share Lower Schooling	0.0041	-0.0179**	0.0038
	(0.0087)	(0.0083)	(0.0045)
Anderson-Rubin (AR) Test Statistic Chi2(2)	14.16***	11.32***	19.74***
Observations	28,778	28,917	28,973
Controls			
State FE	YES	YES	YES
Individual & Household Controls	YES	YES	YES
Geographical Characteristics	YES	YES	YES
Historical Variables	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the District Level. *** p<0.01, ** p<0.05, * p<0.1, + p=0.131. Household Controls: SC (dummy), ST (dummy), Hindu (dummy), Household Perform Livestock, Land Owned, and Land Owned Square. Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Distance from City, Mean Rainfall (annual), Soil Quality (sandy), Mean Temperature (annual), and District Population (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table A4: Agricultural Productivity Mechanism (IV - Technical Advice)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): First stage - Tertiary Educated Member in Household (dummy)								
Mean District Distance Catholic Missionary (log)	-0.0173***	-0.0173***	-0.0173***	-0.0173***	-0.0173***	-0.0173***	-0.0173***	-0.0173***
	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)	(0.0065)
Kleibergen-Paap rk Wald F Statistic	7.113	7.114	7.114	7.114	7.104	7.121	7.095	7.124
Panel (b): Second stage								
	Extension Agent	Krishi Vigyan Kendra (Farm Science Centre)	Agricultural University / College	Private Commercial Agents (including drilling contractor)	Progressive Farmer	Radio / TV / Newspaper / Internet	Veterinary Department	NGO
Tertiary Educated Member in Household (dummy)	0.7874	0.0386	0.0117	0.8748	0.1189	3.6036**	2.6648**	-0.1078
	(0.8303)	(0.3802)	(0.2771)	(0.7915)	(1.2141)	(1.6274)	(1.3163)	(0.2457)
Anderson-Rubin (AR) Test Statistic Chi2(1)	0.98	0.01	0.00	1.25	0.01	10.36***	7.08***	0.20
Weak IV Robust 95% Confidence Interval	[-0.9549, ∞]	[-1.0003, 1.2581]	[-0.8332, 0.7908]	[-0.9116, ∞]	[-3.4866, 3.7243]	[1.4773, ∞]	[0.7363, ∞]	[-1.0516, 0.4468]
Observations	29,595	29,593	29,593	29,593	29,592	29,591	29,590	29,588
Controls								
State FE	YES	YES	YES	YES	YES	YES	YES	YES
Individual & Household Controls	YES	YES	YES	YES	YES	YES	YES	YES
Geographical Characteristics	YES	YES	YES	YES	YES	YES	YES	YES
Historical Variables	YES	YES	YES	YES	YES	YES	YES	YES

Notes: - Robust standard errors are in parentheses and clustered at the District Level. *** p<0.01, ** p<0.05, * p<0.1. Household Controls: SC (dummy), ST (dummy), Hindu (dummy), Household Perform Livestock, Land Owned, and Land Owned Square. Geographical Characteristics: Latitude, Latitude Square, Longitude, Longitude Square, Coastal (dummy), Average River Length, Average Height District, Distance from City, Mean Rainfall (annual), Soil Quality (sandy), Mean Temperature (annual), and District Population (log). Historical Variables: Fraction of Brahman 1931, Fraction of Tribal 1931, Fraction of Urban 1931, Princely State (dummy), and Rail (1909).

Table A5: Variable Descriptions and Sources

	Dependent Variable(s)
Consumption per Capita (log)	Log of one plus village-level mean predicted per capita consumption, estimated using the small-area estimation method Source: SHRUG V-1.5
VIIRS Nightlight per Square km Annual (log)	Log of one plus the ratio of the total light luminosity (measured in radians and ranging from 0 to ∞) in a village to the total area of the village in square kilometres Source: https://eogdata.mines.edu/products/vnl/ and Map of India, processed by Arc GIS
	Independent Variable
Share Tertiary & Above (pop 25+)	The ratio of the total population with education level graduate and above to the total population aged twenty-five and above, multiplied by 100 Source: SECC 2011
Share Below Graduate (pop 05+)	The ratio of the population with some level of education but not a graduate degree to the total population aged 5 years and above, multiplied by 100 Source: SECC 2011
	Instrument Variable
Nearest Distance from Catholic Missionary	To construct the instrument we calculate the minimum distance from the village centroid to the nearest Catholic missionary take the mean of this variable at the subdistrict level and transform it into log Source: Karl Streit (1913), Atlas Hierarchichus, Map of India, processed by Arc GIS
	Current Controls

SC Proportion	The ratio of the total Scheduled Caste (SC) population to the total population of a village plus one Source: SECC 2011
ST Proportion	The ratio of the total Scheduled Tribe (ST) population to the total population of a village plus one Source: SECC 2011
Population (log)	Log of one plus total population of a village Source: SECC 2011
	Geographical Characteristics
Latitude	Latitude coordinates of the centroid of a village Source: Map of India, processed by Arc GIS
Longitude	Longitude coordinates of the centroid of a village Source: Map of India, processed by Arc GIS
Coastal (dummy)	An indicator if a district boundary is on the coast Source: Map of India
Average River Length	Length of rivers that pass through the district in kilometers Source: https://diva-gis.org/
Average Height District	Average elevation of the district in kilometers Source: https://diva-gis.org/
Mean Rainfall (annual)	Mean rainfall is the average of monthly rainfall for the year 2011 - Monthly rainfall is the amount of rainfall in a particular month Source: Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit
Soil Quality (sandy)	The proportion of soil that is sandy in a district Source: 'Soils of India' database of NBSS & LUP, Indian Council of Agricultural Research
Nearest Distance from Big City	The minimum distance from the village centroid to the nearest city with a population of over one million in 2011 - According to the Census of India 2011, there were 53 such cities Source: Census 2011, Map of India, processed using ArcGIS

Mean Temperature (annual)	Mean temperature is the average of monthly temperatures for the year 2011 - Monthly temperature is the average of daily temperatures within a month, with daily temperature defined as the midpoint between the minimum and maximum temperatures Source: Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit
Village Area Square km (log)	Total area of a village in square kilometers Source: Census 2011, Map of India, processed using ArcGIS
	Historical Variables
Fraction of Brahman (1931)	District-level share of Brahman population 1931 Source: Census 1931
Fraction of Tribal (1931)	District-level share of tribal population 1931 Source: Census 1931
Fraction of Urban (1931)	District-level share of urban population in 1931 Source: Census 1931
Princely State (dummy)	An indicator for districts that were in Princely India in 1931 Source: Census 1931
Rail (1909)	An indicator if a railway line passed through a district in 1909 Source: Indian Administrative Report on Railways, map was processed by Arc GIS
	Historical Variables in Table 3
Fraction of Urban (1901)	District-level share of urban population in 1901 Source: Census 1901
Fraction of Brahman (1901)	District-level share of Brahman population in 1901 Source: Census 1901
Fraction of Lower Castes (1901)	District-level share of population recorded as lower caste in 1901 based on each province's enumeration of lower castes Source: Census 1901 and Chaudhary (2009)
Fraction of Tribal (1901)	District-level share of tribal population in 1901 Source: Census 1901

Ethnic Fraction (1901)	A Herfindahl-based measure of caste and religious fractionalisation ranging from 0 to 1 Source: Census 1901 and Chaudhary (2009)
Total Population (1901)	Total population of a district in 1901 Source: Census 1901
Colleges (1901)	Number of colleges in 1901 at district-level Source: District Gazetteers of India
School (1901)	Number of schools in 1901 at district-level Source: District Gazetteers of India
Income Tax per Capita (1901)	Income tax revenues per capita in 1901 at district-level Source: District Gazetteers of India
Variables Used for Robustness in Table 9	
Under 5 Mortality Rate	Under 5 Mortality Rates of a district in 2011 - Link Source: Census, 2011
Share Irrigated Land	District-level share of land that is irrigated Source: Census, 2001
Violent Crimes per Capita	Total Violent Crimes per 10,000 population in 2004 at district-level - Violent crimes include murder, attempt to commit murder, culpable homicide not amounting to murder, attempt to commit culpable homicide deaths, kidnapping and abduction Source: National Crime Record Bureau, India, 2004
Confidence Panchayats	District-level share of households that report that they have a great deal of confidence in village panchayats/nagarpalika (local councils) to implement public projects Source: Indian Human Development Survey, 2004-05
Agricultural Yield, 1956 (Log)	District-level agricultural yield in 1956 transformed into log. Source: Banerjee and Iyer (2005)
Rural Migration Rate	Share of rural individuals aged 25 and above who migrated within the last one year. Source: National Sample Survey 2007-08 (64th Round)

School Index	An indicator variable based on the median value of the principal component, constructed using the indicator variables of primary, middle, secondary, and higher secondary schools in a village Source: Census 2011
Primary Health Center	An indicator for a Primary Health Center in a village Source: Census 2011
Maternity Center	An indicator for a Maternity Center in a village Source: Census 2011
Post Office	An indicator for a Post Office in a village Source: Census 2011
Drainage System	An indicator for Drainage in a village Source: Census 2011
Variables Used for Robustness in Table 10	
District Urbanization Rate	The ratio of the total urban population to the total population of a district, multiplied by 100 Census 2011
Subdistrict Urbanization Rate	The ratio of the total urban population to the total population of a sub-district, multiplied by 100 Census 2011
Relative Urban Wage Tertiary vs. Non-Tertiary Education	The ratio of district-level urban tertiary-educated wages to district-level urban non-tertiary-educated wages for individuals aged 25 and above Source: National Sample Survey – Employment and Unemployment, 2009–2010, 66th Round
Labour Force Participation Tertiary Educated (urban)	The ratio of employed urban tertiary-educated individuals to the total urban tertiary-educated population at the district level for individuals aged 25 and above Source: National Sample Survey – Employment and Unemployment, 2009–2010, 66th Round
Variables Used for Mechanism in Table 11	

Tertiary Educated Member in Household (dummy)	An indicator equal to 1 if at least one household member above age 25 has completed tertiary education, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013
Share Lower Schooling	The share of household members aged 25 years and above who have attained some level of education but do not hold a graduate degree, relative to the total household members aged 25 years and above, multiplied by 100 Source: Situation Assessment Survey of Agricultural Households, 2013
Share Below Graduate (pop 05+)	The ratio of the population with some level of education but not a graduate degree to the total population aged 5 years and above, multiplied by 100 Source: SECC 2011
Agriculture Output per Unit of Land	The total value of agricultural output produced by the household divided by the total land cultivated (in INR per hectare) Source: Situation Assessment Survey of Agricultural Households, 2013
Access to Technical Advice	An indicator equal to 1 if the household has access to technical advice related to agriculture, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013
Crop Diversification	The share of crops grown by the household other than cereals and pulses Source: Situation Assessment Survey of Agricultural Households, 2013
SC (dummy)	An indicator equal to 1 if the household belongs to the Scheduled Caste (SC) social group, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013

ST (dummy)	An indicator equal to 1 if the household belongs to the Scheduled Tribe (ST) social group, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013
Hindu (dummy)	An indicator equal to 1 if the household reports Hinduism as its religion, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013
Household Perform Livestock	An indicator equal to 1 if the household engages in livestock rearing activities, 0 otherwise Source: Situation Assessment Survey of Agricultural Households, 2013
Land Owned	Total land owned by the household (in hectares) Source: Situation Assessment Survey of Agricultural Households, 2013
Variables Used for Mechanism in Table 12	
Net Primary Productivity (log)	Log of one plus the satellite-based estimates of agricultural yields, which are a common unit of productivity across different crop types in designated cropland areas Source: European Space Agency (ESA) satellite database, Map of India, processed using ArcGIS
Variables Used for Heterogeneity in Table 13	
Good Soil Quality	An indicator for below-median 'Soil Quality (sandy)' in a district Source: 'Soils of India' database of NBSS & LUP, Indian Council of Agricultural Research
Bad Soil Quality	An indicator for above-median 'Soil Quality (sandy)' in a district Source: 'Soils of India' database of NBSS & LUP, Indian Council of Agricultural Research

Bad Rainfall	An indicator for below-median 'Mean Rainfall (annual)' in a village Source: Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit
Good Rainfall	An indicator for above-median 'Mean Rainfall (annual)' in a village Source: Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit
Variable Used in Table Table 14	
Share Non-Farm Employment 2011	Share of workers engaged in non-farm activities classified as other workers excluding cultivators, agricultural labourers and household industry workers in 2011 Source: Census of India, 2011
Variables Used for Mechanism in Table 15	
Tertiary Educated (dummy)	An indicator for an individual whose age is above 25 and completes tertiary education Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Regular Wage Earner	An indicator for an individual whose principal activity status is as a regular salaried/wage employee Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Private Sector	An indicator for an individual employed in the private sector Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Government Sector	An indicator for an individual employed in the government/public sector Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round

Urban Work Location	An indicator for an individual whose place of work is located in an urban area Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Gender	A binary indicator equal to 1 if the individual is male and 0 if female Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Age	The age of the individual in completed years Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
SC (dummy)	An indicator equal to 1 if the individual belongs to the Scheduled Caste (SC) social group, 0 otherwise Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
ST (dummy)	An indicator equal to 1 if the individual belongs to the Scheduled Tribe (ST) social group, 0 otherwise Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Hindu (dummy)	An indicator equal to 1 if the individual reports Hinduism as their religion, 0 otherwise Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Land Owned	Total land owned by the household (in hectares) Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Household Size	Total number of members in the household Source: National Sample Survey - Employment and Unemployment, 2009-2010, 66th Round
Instrument Variable Used in Appendix Table Table A1 and Table A3	

Share Primary Schooling 1991	Share of population that completed only primary education in 1991 Source: Census of India, 1991