Transforming Rural Economies Through Tertiary Education: Evidence from India *

Amparo Castelló-Climent† Abhiroop Mukhopadhyay‡ Ravinder‡

† *University of Valencia*

‡ *Indian Statistical Institute, Delhi*

February 2025

Abstract

This paper analyses the role of tertiary-educated in rural development and shows that skilled workers have had an important impact on rural prosperity. Using census data in India, we find that a one percent increase in the share of village population with tertiary education raises per capita consumption by around 7.2 percent. The results hold with night-light density data and are not driven by lower levels of education, omitted variables, or atypical observations. Skilled workers affect the village economy by promoting agricultural productivity and engaging in productive private sector jobs available in the village and nearby commutable areas.

Keywords: Rural, Development, Higher Education.

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

*We would like to thank the participants at the 17th Annual Conference on Economic Growth and Development, 27th ACCDE at Jadavpur University, 2023 CSH-AXA International Conference on Socio-Economic Inequality in India, 2022 Advanced Graduate Workshop at Azim Premji University, and seminar participants at Warwick University, Indian Institute of Technology in 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 while Ravinder was a visiting Ph.D. fellow at UNU-WIDER in Helsinki, and he is thankful to Professor Kunal Sen for his guidance and mentorship during this period.

1 Introduction

Economic development is associated with a process of structural transformation in which the demand for workers outside the agricultural sector increases. The movement of labor from agriculture to manufacturing and services can be accelerated by an increase in the supply of educated workers, as human capital is more valuable outside agriculture (Porzio et al., 2022). Typically, such a movement comes with a transfer of labor from rural to urban areas (Lewis et al., 1954; Harris and Todaro, 1970; Gennaioli et al., 2013). However, in less developed countries, despite the increase in education levels in the last decades, most of the population still resides in rural areas. This is particularly relevant in the case of India, where approximately one-third of tertiary educated individuals resided in rural areas. In this paper, we analyze the role of tertiary-educated on rural development and show that skilled workers have had an important impact on rural prosperity. We find that a one percentage point increase in the share of village population with tertiary education raises per capita consumption by around 7.2 percent.

Human capital has played an important role in promoting regional development in India. Tertiary education is closely related to better prosperity at the state level and at the district level (Castelló-Climent and Mukhopadhyay, 2013 and Castelló-Climent et al., 2018). The mechanism at work is through the employment of skilled workers in the services sector, which is consistent with a structural transformation in which higher education leads people to move from the least productive to the most productive sectors. Typically, such movements come with a transfer of labor from rural to urban areas (Lewis, 1954; Harris-Todaro, 1970; Gennaioli et al. (2013)). This process leads to a positive link between an increase in human capital, a higher share in the services sector, and an increase in urbanization, which in turn leads to growth and development (Sen, 2016; McMillan et al., 2017; Swiecki, 2017; Herrendorf and Schoellman, 2018).

Whereas the movement of highly educated workers to urban areas and their influence on different aspects of development has been analyzed by the literature, little is known about the influence of highly educated individuals on rural development, where the skilled labor force has grown rapidly over time. In India, according to the 2011 Census, approximately one-third of tertiary-educated people reside in rural areas. Figure 1 shows a tremendous increase in recent decades. For the age group of 25-34, the share of tertiary educated people living in rural areas almost doubled, increasing from 6.91 percent in 2011 to 12.29 percent in 2018. What are the implications of such an increase for rural development? Does tertiary education lead to an increase in rural prosperity? If so, what are the mechanisms?

In this paper, we estimate the impact of a higher share of tertiary-educated in the village labor force on village prosperity in around 427,000 villages in 27 states of India. Estimating the effect of education on development at the village level, which is considered the smallest spatial unit in India, allows us to incorporate externalities at the local level, in line with the evidence of college externalities in cities (Moretti, 2004).[1]

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. The advantage of our empirical strategy is that we use intra-state variation. 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. However, since the skills of decision makers at the local level could affect development (Casey et al., 2023), we control for the provision of public goods at the village level. We also account for other potential omitted variables bias by controlling for a broad set of current demographic controls, geographical characteristics, and historical variables.

Any remaining threats to the exogeneity of tertiary education are addressed with an instrumental variable. We identify the impact of tertiary education using the variation in the historical location of the Catholic mission in India circa 1911 (Castelló-Climent et al., 2018). Catholic missions changed the supply and preference for higher education in India due to their emphasis on providing quality education. 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 district 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 was on the coast, whether the district had a railway line passing through, and whether the district was tribal. Most importantly, the

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

locations were not determined by the income of the areas, as predicted by income tax collections (Castelló-Climent et al., 2018). Hence, such a location, conditional on controlling for the predictors, is likely to be exogenous.

One of the features of catholic missionary locations was that Catholic missionaries settled in towns and cities while conducting their activities in the rural areas around them. Villages close to towns would then be close to catholic missions. Therefore, we use the mean distance to the nearest Catholic missionary, averaged over all villages in a sub-district as our instrument (for the ease of presentation, we refer to it henceforth as the mean distance from Catholic mission). Our measure of location has predictors similar to determinants of district location of Catholic missions. Most importantly, it does not correlate with historic income tax collections, though it does correlate with historic urbanization rates.[2] To remove any impact due to urbanization, we control the distance from a village to the nearest city and the historical urbanization rate in all our regressions.

Our first stage results show that the mean distance from the Catholic mission is highly correlated with the rural share of tertiary completion. The closer the average village in a sub-district is to the historical location of the Catholic mission, the higher the tertiary completion. We provide several test for identification. The Kliebergen-Paap F-statistic, the Anderson-Rubin test, and IV robust 95 percent confidence intervals show the instrument is not weak. We also document that the instrument meets the exclusion restriction. We show that the instrument does not work through alternative channels different from that through tertiary education.

Using this exogenous source of variation for the current share of university-educated individuals, we find that, on average, villages with a higher proportion of college-educated individuals have higher prosperity, measured by village-level mean per capita consumption taken from the Socioeconomic High Resolution Rural-Urban Geographic Data Platform for India (SHRUG). The results hold with a wide range of controls, including lower levels of education. The quantitative effect is non-negligible: a one percentage point increase in the share of the population with tertiary education increases log mean consumption per capita by 7.2 percent.

We show the results are robust to a variety of sensitivity tests. In the first place, we

[2] This correlation is somewhat mechanical as the larger the urban part of the district, the closer the rural part of the district is to it.

proxy development with a measure of consumption per capita, as there is no data available on income per capita at the village level. This measure is imputed consumption, rather than actual consumption. To check the robustness of our results, we use 1000 replications of bootstrapping log consumption per capita and show the results are not driven by measurement error. We also show the results are robust to alternative measures of income, such as night light density data, which have been proven to be a good proxy for development (Henderson et al., 2012, Micholapoulos and Papaioannou, 2013, 2014; Alesina et. al, 2012). We use the Visible Infrared Imaging Radiometer Suite (VIIRS) Night Lights (2015), extracted at the village level, and show the results hold.

The share of tertiary educated individuals contains a large number of zeros, generating a skewed distribution. We show the results hold when we transform the variable using an inverse hyperbolic sine function (Bellemare and Wichman, 2020). We also document that our results are not driven by outliers, atypical observations, or extreme values.

Missionaries were active in other aspects of the village that could have affected development, such as health, infrastructure, or institutions. We show that our instrument is not correlated with measures of infant mortality, violent crime rates, confidence in local village councils, district irrigation coverage, and several measures of village infrastructure, including primary health centers and maternity centers. Moreover, the inclusion of these variables in the regression does not influence the impact of education on village development.

To rule out that our results are driven by the fact that missionaries settled in urban centers and could pick up the effect of urbanization, we also show that our results remain unchanged when we account for the subdistrict urbanization rate or for urban wages in the district.

The effect of tertiary education on village prosperity can be driven by several factors. The limited available data at the village level force us to analyze only a few of the plausible mechanisms. Given that rural economies are largely based on agriculture, we first analyze whether tertiary-educated workers have had an impact on agricultural productivity. A major concern is that there are no available village-level data on crop yields. To overcome this shortcoming, we use two data sets to proxy agricultural productivity. The first data set contains microdata on household agricultural profits, taken from the Nationally representative survey of agricultural households (Situation Assessment Survey of Agricultural Households, collected by the National Sample Survey Organization (NSSO), 2013). We examine the impact of having a tertiary-educated member in the household on several measures, including the gross agricultural value of crops per unit of land cultivated, crop diversification, and access to technical advice regarding agriculture. We find that households with tertiary educated members show a higher agricultural gross value per unit of land. They are also more likely to diversify their crops (Hazra, 2001; Joshi et al., 2004; Birthal et al., 2015), and have better access to technical knowledge related to agriculture, suggesting that tertiary educated individuals play a positive role in the adoption of new and more productive technologies. We corroborate these findings with satellite-based spatial Net Primary Productivity (NPP), as a measure of agricultural productivity (Zaveri et al., 2020). The results indicate that a higher share of tertiary education leads to higher agricultural productivity.

Tertiary education can also influence development through channels outside agriculture. Ideally, we would like to decompose the income in the village into its constituent elements to see what drives the impact of tertiary education. As mentioned above, we are unable to perform a full decomposition due to the lack of a representative data set at the village level. Instead, we provide some suggestive evidence on the impact of tertiary-educated on the probability of being in skilled occupations, characterized by jobs that are productive and pay well. To this aim, we rely on micro-data from the Employment-Unemployment Survey conducted by the NSSO in 2009. The results of rural India indicate that, compared to less educated workers, tertiary educated individuals have a higher probability of working in regular wage jobs in the private sector. In contrast to the public sector, private wages reflect the value of the marginal product, so tertiary education not only leads to higher income but also to higher productivity.

An alternative mechanism could be driven by tertiary educated daily commuting to the nearest town. While we show that urbanization does not confound our results, some of the non-farm private jobs are likely to be in nearby towns. Not surprisingly, the results indicate that tertiary education also increases the probability of commuting to urban areas and working as a regular wage earner in the private sector. This channel, however, is unlikely to be the main source of village prosperity, as the proportion of tertiary educated people who commute daily to the nearest urban center is only 9 percent, which suggests that a significant part of the prosperity is generated within the village.

This suggests a model of structural transformation that does not rely on moving educated people through migration to urban areas. Instead, this relies on tertiary-educated people living in the village, where a few of them commute to nearby towns, while others work in skilled occupations in the private sector. Interestingly, even those who do not work in the agriculture sector have a positive impact on agricultural productivity through externalities at the household level.

Our paper is related to the literature that points to highly educated individuals as drivers of development and growth.[3] The emphasis on tertiary education has been particularly important in India. High-quality engineering and technology-oriented higher education institutions have been the aim of all Indian governments over the years. As a result, the share of the population with tertiary education is higher in India than in China. In 2015 the share of the working-age population (15-64 years) with tertiary education was 12.68 percent in India compared to 7.27 percent in China. The share of college graduates has translated into higher growth and development in the states and in the districts of India (Castelló-Climent and Mukhopadhyay, 2013; Castelló-Climent et al., 2018). In this paper, we reinforce this literature by showing that in addition to their impact through the service sector in urban areas, college graduates have also had an important impact on rural development in the villages of India.

This paper is also related to the literature on structural transformation. As countries become richer, there is a reallocation of workers from agriculture to manufacturing and services (McMillan et al., 2017). The average wages are lower in agriculture than in the other sectors (Herrendorf and Schoellman, 2018). One of the explanations is that agriculture workers have relatively lower human capital than other workers (Caselli and Coleman II, 2001).[4] Higher education leads individuals to move from the agriculture sector to manufacturing and service. Typically, such a movement comes with the migration of labor from rural to urban areas (Lewis et al., 1954; Harris and Todaro, 1970; Gennaioli et al., 2013). One of the motivations behind urban migration is the higher wage in urban areas (Baum-Snow and Pavan, 2012; Roca and Puga, 2017), which is particularly relevant for higher-educated individuals as higher-level

[3]Historical evidence shows that skilled workers played a key role in the Industrial Revolution (Meisenzahl and Mokyr, 2011; Squicciarini and Voigtländer, 2015; Maloney and Valencia Caicedo, 2017). Using contemporary data, Gennaioli et al. (2013) find that managerial education plays a larger role in explaining regional development than human capital of workers.

[4]The assumption that non-agriculture is more human capital intensive than agriculture is consistent with widely documented patterns of sorting of high-skilled workers in non-agriculture (e.g., Gollin et al., 2014; Young, 2013; Porzio and Santangelo, 2017), larger returns to skills in non-agriculture (Herrendorf and Schoellman, 2018) and skill-specific mobility across sectors (Hicks et al., 2017).

specialized jobs are concentrated in cities. We complement this evidence with a structural transformation also driven by skilled workers, but that occurs in a less developed country within rural areas and within the agricultural sector.

The evidence is consistent with previous findings that show that in the Indian context, human capital is an important factor that facilitates the adoption of new and more productive technologies (Foster and Rosenzweig, 1995, 1996). However, in contrast to previous studies that highlighted the importance of primary education during the era of the Green Revolution as a background (Schultz (1975)), we use a more recent context where agriculture is less traditional and rural tertiary completion rates are five times higher. Interestingly, some of our effects also originate from having tertiary educated members in the household, even though they may not work in agriculture. It is, therefore, plausible that they are sources of information for new technology and markets.

The structure of the article is as follows. In section 2, we describe the data and set out the empirical methodology. In section 3 we present the main results. 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 Data and Empirical Strategy

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.[5] The analysis focuses on the share of the population that has completed tertiary education, that is, those with a university degree/diploma and above. We source these data from the Socio-Economic and Caste Census (SECC), collected in 2011.[6] Precisely, we define *Share_ter* as the share of the population who have completed tertiary education. To ensure

[5] 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.

[6] 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.

that our measure is not right censored by population who are not old enough to reach their highest education level, we use the population that is 25 and above as the relevant population.[7] Figure 2 shows the spatial distribution of the share of tertiary education across subdistricts in rural India. 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).[8] As is standard in the literature, we use the log of mean consumption per capita as our dependent variable.[9]

In our data, the mean value of log consumption per capita is 9.66, and that of the share of tertiary education is 5.63 percent.[10] The unconditional correlation between both variables 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 historical

[7]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.

[8]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 article.

[9]Since consumption is imputed, it has a distribution. For our analysis, we use 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.

[10]The share of tertiary educated may suffer from some measurement error as the maximum value is 100 %. The mean for the top 1% is 27.2 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%.

controls in our model, and estimate the following specification:

$$ln\_Cons\_per\_Capita_{vds} = \alpha_s + \theta Share\_Ter_{vds} + \rho^{'} C_{vds} + \pi^{'} G_{vds} + \delta^{'} H_{ds} + \epsilon \quad (1)$$

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.[11]

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), sourced from http://www.diva-gis.org/ and extracted using Arc GIS tools; annual rainfall and temperature in a village, taken from the 'Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit;' [12] the area of the village, from the Census of India 2011; 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

[11]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.

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

LUP, Indian Council of Agricultural Research; and the distance to the nearest million plus city calculated from each village,[13] using Arc GIS.

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.[14,15] These variables account for the evolution of other contemporaneous variables. The summary statistics for the main variables are reported in Table 1.

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. To address this challenge, we analyze the impact of higher education using the variation in the historic location of Catholic missions in India circa 1911. As explained in Castelló-Climent et al. (2018), 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. 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).[16]

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

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

[14] 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.

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

[16] 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).

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. (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 an 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 district location, 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. However, understandably, the missionaries settled in cities and towns within districts, while they worked in rural areas around the 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.

3 Main Results

We start with a descriptive 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, as measured by per capita consumption. However, the finding could be driven by omitted variables related to both tertiary educated individuals and consumption per capita. Hence, in columns (2-5) we control for state fixed effect and current, geographical, and historical variables. The results show a positive and statistically significant coefficient for tertiary education even after including a wide range of controls. Another concern is that college-educated individuals could be 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 (see Table A1 in the Appendix).[17] Thus, there is a strong correlation between the proportion of adult individuals with a college education and greater economic prosperity in Indian villages.

The previous results mainly show the correlation among the variables. To make causal inference, 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). The results indicate that villages located farther away from a Catholic mission in 1911 have a lower share of the population with tertiary education in 2011. Although the Kliebergen-Paap F stats are high to begin with, they fall as we add geographical and historical controls. Hence, we report Anderson-Rubin Test statistics for weak instrument as well as report Weak IV robust 95% confidence interval in panel (b).

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). The coefficient with the full set of controls (column 5) implies that a one percentage point increase in the share of tertiary educated increases per capita consumption by 7.2 percent.

As noted above, it could be argued that the share of tertiary education is picking up the impact of other levels of schooling. In Table A1 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

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

$$Share\_Below\_Graduate\_(pop05+) = \frac{Total\_Pop\_with\_some\_level\_of\_education\_but\_not\_graduate}{Total\_Pop\_age\_(05+)} * 100$$

As mentioned earlier, we don't have an age-specific distribution of education levels. We estimate the share of 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).

education.[18]

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.

4 Robustness

In this section, we address various concerns regarding our estimates and provide various sensitivity checks.

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, 1986), utilizing 1000 draws of log consumption per capita provided by the SHRUG data set. As depicted in Figure 3, the distribution of the coefficient remains positive and exhibits a range between 0.059 and 0.095. The mean and median of the distribution are similar: around 0.073. 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.

[18] Our strategy to control for the lower levels of education only shows that the effect that we pick up as the effect of tertiary education is not driven by lower levels of education. But we would hasten to add that without an instrument for lower levels of education, we do not make any claim on the impacts of lower levels of education

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.[19]

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 all 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.[20] 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 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

[19]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.

[20]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.

(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. 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. 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.[21]

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

[21] The source for all variables, except one, is the census; confidence in local political institutions is a district level average created 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 of India.

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. Despite the impressive growth of the service sector, India is still an agricultural society. In 2011, 54.6 percent of the workforce were employed in agriculture. Although the share of Indians working in agriculture is declining, it remains the main sector of employment. Thus, we analyze whether tertiary education has influenced development through the agricultural sector. Then we document the effect of tertiary education on skilled occupations, providing some insight on the probability that high-skilled workers commute to the nearest towns on a daily basis.

5.1 Agriculture

The existing literature on the impact of human capital on agricultural productivity is ambiguous. While the early literature argued that education was important for agricultural development (Schultz, 1964), recent studies provide mixed effects. Some studies do not find any effect, or even a negative impact of human capital (Craig et al., 1997; Vollrath, 2007), while others report strong positive effects (Foster and Rosenzweig, 1996; Reimers and Klasen, 2013). Such positive impacts are mediated through the enhancement of decision-making skills, the access to information on inputs and prices, the faster adoption of promising new technologies, and general preferences for riskier production technology that typically promises higher returns (Asadullah and Rahman, 2009).

To explore whether tertiary education has an impact on agricultural income and the various mechanisms highlighted in the literature, we use data collected in 2013 by the National Sample Survey Organisation(NSSO), as a part of the "India-Situation Assessment Survey of Agricultural Households."[22] The data includes 29,652 households that are involved in cultiva-

[22] There is no large representative village level data on agriculture incomes or agricultural practices, which is

tion. We use three proxies for agricultural productivity. Following Deolalikar (1981), we first measure agricultural income by the gross value of agricultural produce per unit of cultivated land.[23]. Second, the literature has found robust evidence on the potential of crop diversification to reduce the downside risks of farming (Di Falco and Chavas, 2009; Di Falco et al., 2011; Bezabih and Sarr, 2012; Bozzola et al., 2018; Bozzola and Smale, 2020). To account for crop diversification, given that almost all farmers grow cereals, we use the share of crops cultivated by the household that are neither cereal crops nor pulses. Furthermore, while the survey does not have detailed information on the technology adopted, it provides information on whether households have access to technical advice regarding agriculture. We use access to such advice as a proxy for access to information about technology.

The gross value of crops, the amount of land cultivated, and the access to technical advice are measured at the household level. However, the relevant level of education is measured at the individual level. This requires the information on education to be aggregated at the household level as well. To that end, our independent variable is a dummy variable that indicates the presence of a tertiary educated member in the household. As a result, we look at the impact of tertiary educated workers who work in agriculture, as well as those who may not be working in agriculture, but may have an effect on it by providing information to the household related to markets, technology, and prices.[24]

Our instrument also has to be adjusted, as in the agriculture survey data the lowest level of geographic unit that can be identified is the district. To use an instrument analogous to the one used in the village regressions, we use the log of the average distance to the catholic missionary averaged across villages in the district. Given that now the instrument varies at the district

why we move to household level outcomes. According to Reimers and Klasen (2013), some of the inconsistencies in the literature between micro and macro results with respect to the impact of education on agricultural productivity is due to the choice of wrong education variables such as enrollment and literacy; if one uses education attainment instead, then both micro and macro returns to education are positive. This is consistent with the variable we use to measure human capital.

[23]We do not use crop yields as the numerator because multiple crops are grown by each household and the choice of crop is endogenous. We also do not calculate the value of output net of all input costs as imputing value of unpaid family labour requires a lot of assumptions on the opportunity cost of labour. Analogously, since we do not know how much labor time goes into cultivation, we cannot calculate the agricultural output per worker

[24]In the sample, 24 % of tertiary educated individuals are classified as "market-oriented skilled agriculture and fishery" workers

level, we cluster our standard errors at the district level. Moreover, in addition to the district-level control variables used above, we include a broad range of household-specific controls, such as the mean age of the household, dummy variables for belonging to disadvantaged social group (Scheduled castes and Scheduled tribes), being a Hindu household, the household size, the number of males, the number of children, the land owned and its square, and whether the household performs livestock activities, since this has implications on land usage.

The results, reported in Table 11, show that in agricultural households, having a tertiary-educated individual positively impacts the gross income per acre of land (column 1). Households with tertiary educated members also tend to have greater access to technical advice (column 2) and greater diversity in their choice of crops (column 3).[25] Overall, these results indicate that having a tertiary-educated member in the household has a positive impact on agricultural productivity.

Nevertheless, survey-based data on gross agricultural income at the micro-level can be subject to measurement error. To verify the robustness of the findings, we complement the information with satellite-based data, which also allows us to confirm whether the results hold at a more aggregate level. We use data on Net Primary Product (NPP), which is a satellite-based measure of agriculture productivity (Zaveri et al., 2020). The data is gridded, and each grid covers many villages. We aggregate the data to calculate the agriculture productivity for each subdistrict. The average rural tertiary completion rate, the log of average distance to the catholic missionary, and the current controls are also calculated at the subdistrict level. The results in Table 12 confirm the micro findings above; subdistricts with higher rural tertiary completion rates have a positive causal association with agricultural productivity, measured by Net Primary Product.

The impact of tertiary education on village prosperity is also confirmed by heterogeneity analysis, where we look at the interaction of tertiary education and favorable climate and soil conditions for agriculture. To do so, in columns (1) and (2) of Table 14, we consider the districts where the soil is good and of poor quality, respectively.[26] Interestingly, we see that

[25]The first-stage test for weak instruments show the F-stat is low. To complement this information, we also report the results of the Anderson Rubin test, and the weak instrument robust Anderson-Rubin (AR) confidence intervals. The AR test statistic rejects a weak instrument problem, and the 95 % AR weak instrument robust confidence intervals in each IV regression suggest that the coefficient of interest in each case is bounded away from zero.

[26]We define a district to have good quality soil if the proportion of sandy soil is less than the median proportion

the effect of tertiary education on consumption per capita is statistically significant only for districts where soil quality is better. Likewise, when we conduct an analogous analysis with rainfall, we find that the results are significant only when the district-level rainfall is good. This result reinforces previous findings and indicates that agriculture has a key role to play in how higher education affects village prosperity.

The results are novel and complement the findings in the literature. Most of the seminal papers documenting the role of human capital in agricultural productivity in India are in the context of the Green Revolution (Foster and Rosenzweig, 1996). They find that the returns to agriculture from human capital are high for primary and secondary education. However, in the 1970s and 1980s, tertiary education completion rates were less than 1 and 2 percent, respectively, and were largely concentrated in urban areas. In the 2000s, mechanization in the agricultural sector increased, and information about new technologies has become of upmost importance. The share of tertiary educated in rural areas has also increased dramatically, reaching up to 12 percent of the population aged 25 to 34 in 2018.

In the new context, we show that tertiary-educated individuals play a relevant role in promoting agricultural productivity at the village level. Whereas most of the articles look at the impact of farmers' human capital on productivity, we show that even having a tertiary educated person in the household matters for agricultural productivity. Higher-educated members are more likely to be aware of external markets, prices, and newer technologies. Thus, the results point to a possible new channel through which human capital externalities can impact agriculture productivity.

5.2 Skilled Occupations

The stylized model of human capital and sectoral growth posits that if the non-agricultural sector is more skill intensive than the agricultural sector, then a rise in education among the adult labor force would re-allocate labour from the farm to the non-agricultural sector (Caselli and Coleman II, 2001) and lead to a rise in income. However, in most countries of the world, this changing sectoral allocation has been accompanied by rural out-migration of the educated (Taylor and Martin, 2001). Since we look at the tertiary educated living within the village,

of sandy soil for the country (around 2 percent). When the proportion of sandy soil is the median proportion of sandy soil, we label it as a poor soil quality district the rise of prosperity hints at other processes at play.

To delve deeper, we look at the occupation profile of the tertiary-educated. Our analysis relies on employment data collected by the NSSO over 12 months spanning from 2009 to 2010. We focus on 125,023 adults aged 25 and older living in rural India. Occupation categories are classified into farmers, those running non-farm businesses, also called the self-employed, those working as agriculture labor, those working on daily wages in the non-farm sector, and those with regular wage jobs in the non-farm sector. Although the income profile of all occupations is not available in a representative survey, there is consensus that, on average, a regular wage job, also called a salaried job, is typically the highest paid profession (Fulford, 2014). However, it should be noted that in developing countries, an inefficient public sector contributes significantly to regular jobs and the wage is not necessarily associated with productivity (Pritchett, 2001).

Given that 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 as the instrument the log of distance to the catholic missionary, averaged over villages in a district.[27] Panel (a) of Table 13 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 13, 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. In addition, in line with our previous findings that have shown positive impacts of having tertiary-educated members in the household on agricultural productivity, we use data

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

on occupational categories to check whether having tertiary education has a causal impact on being a skilled farmer or worker in agriculture. The results in column (3) find that this is indeed the case, reinforcing the effect of tertiary education on the agricultural sector.

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 to commute 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 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 village or in the nearest town.[28]

The results suggest an interesting perspective on how we think about structural transformation. The impact of tertiary education on village incomes does not come from migration and remittances. Instead, it comes from commuting, where the tertiary educated reside in villages but commute for work to urban areas. This mechanism, however, is only part of the effect of tertiary education on village prosperity, as the proportion of tertiary educated people who commute daily to the nearest urban center is only 9 percent, suggesting that a significant part of the prosperity is generated within the village.

6 Conclusion and Discussion

The vast majority of the world lives in rural areas, especially in less developed countries. What are the factors that promote development in these areas is of utmost importance. In this paper, our focus is on the role of tertiary education in rural development.

In some countries like India, tertiary education is growing at a high rate. The growth is happening not only in urban places, but also rural areas have witnessed an increase in highly qualified population in the last decades. The literature has focused on the relationship between

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

the increase in tertiary educated individuals, the growth of the service sector, and the increase in urbanization rates, but little is known about whether tertiary educated individuals have promoted rural development and, if so, through which channels this has happened.

The novelty of our paper is the focus on the role of tertiary-educated individuals in rural prosperity. Using several measures that proxy for rural development, the results show a strong effect of a higher share of tertiary educated individuals at the village level on village consumption per capita and night light density. A one percent increase in the proportion of population with tertiary education increases log mean consumption per capita by 7 percent. We show that the results are not driven by lower levels of education, omitted variables, or atypical observations. We also show that our results are not likely to be confounded by institutions, health interventions, or infrastructure.

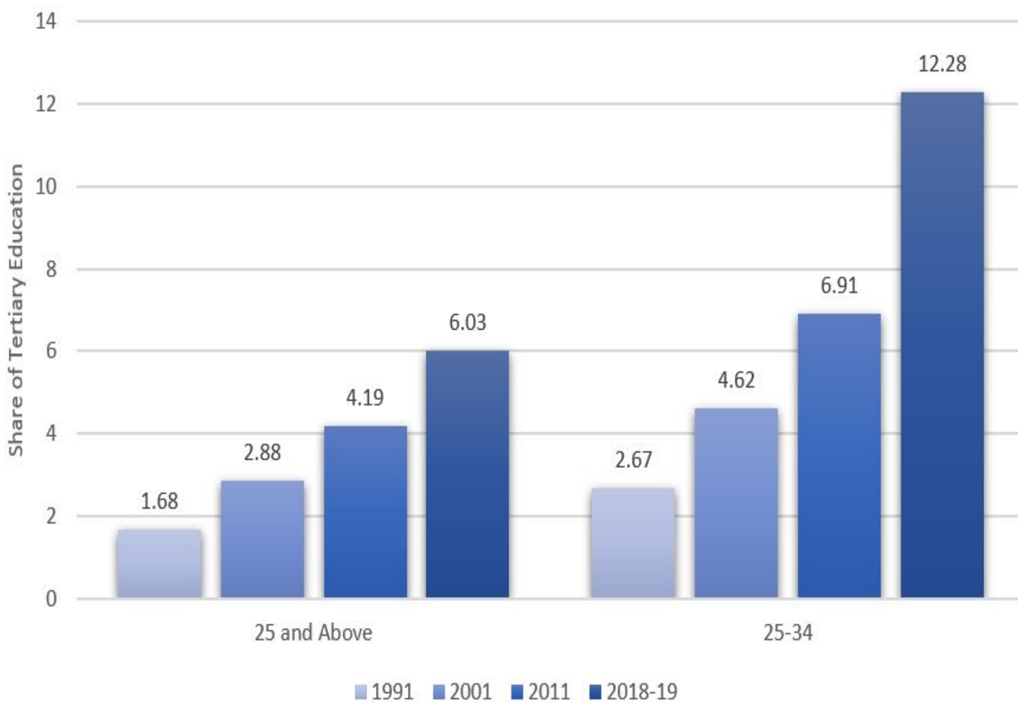
Delving into the mechanisms driving the link between tertiary education and village prosperity, we find an important role for agriculture. The presence of tertiary educated in a household increases agricultural gross income, increases crop diversification, and the household is more likely to have access to technical advice on agriculture. These results are confirmed by satellite data that measures Net Primary Product of an area. They are also supported by the fact that the results are more robust for districts with good-quality soil and rainfall. Furthermore, when we look at the impact of tertiary education on occupations, we find that tertiary education allows adults to work as a skilled farmer/worker in agriculture.

We also find suggestive evidence that tertiary education increases the probability of getting regular jobs in the private sector, as opposed to working on daily wages in the non-farm sector or in the public sector. In less developed countries, regular private jobs are better markers of productive jobs than those in the public sector. Typically, regular jobs or salaried jobs are the highest-paid professions, and the salary could be an indication of the marginal productivity of the worker. The fact that tertiary educated people are more likely to work in the private sector, rather than in the public sector, suggests that higher income is likely to be accompanied by higher productivity.

The stylized models of structural transformation associate higher education with migration to cities to take up non-farm jobs. Also, higher education is often associated with increased urbanization. In this context, village prosperity can be achieved through remittance transfers. However, our results tell a different story. In India, the share of tertiary educated workers residing in rural areas, which has almost doubled in recent years, is associated with an increase in income generated primarily within the village. Tertiary-educated affect the village economy by bringing their skills and knowledge of markets and technology to play in agriculture, and also by engaging in productive private sector jobs available in the village and in nearby commutable areas. A weakness of our analysis, however, is that we have shown the occupation profile, but have not been able to fully decompose the village prosperity into its constituent parts.

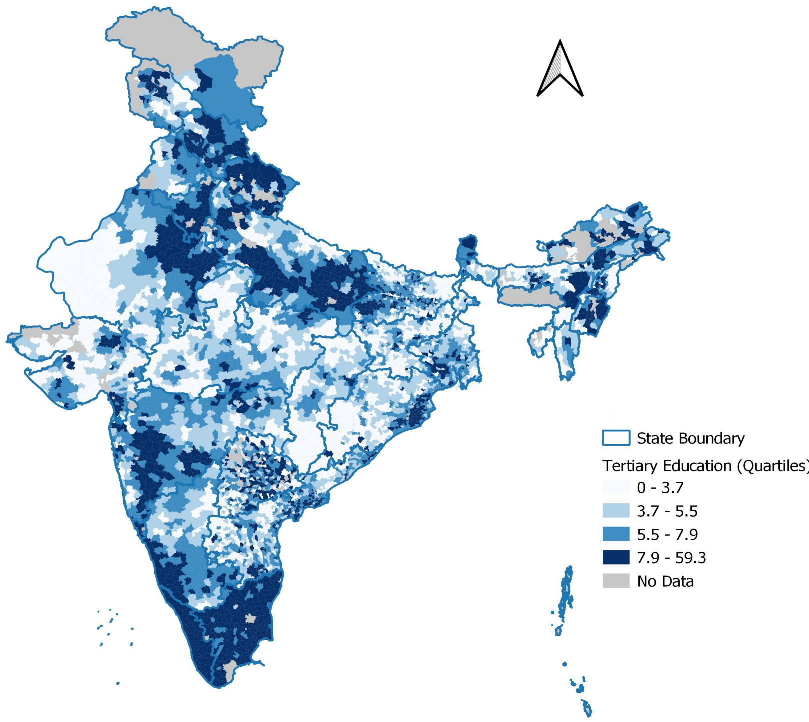
7 Tables & Figures

Figure 1: Share of Tertiary Education in Rural India from the Year 1991 to 2018-19



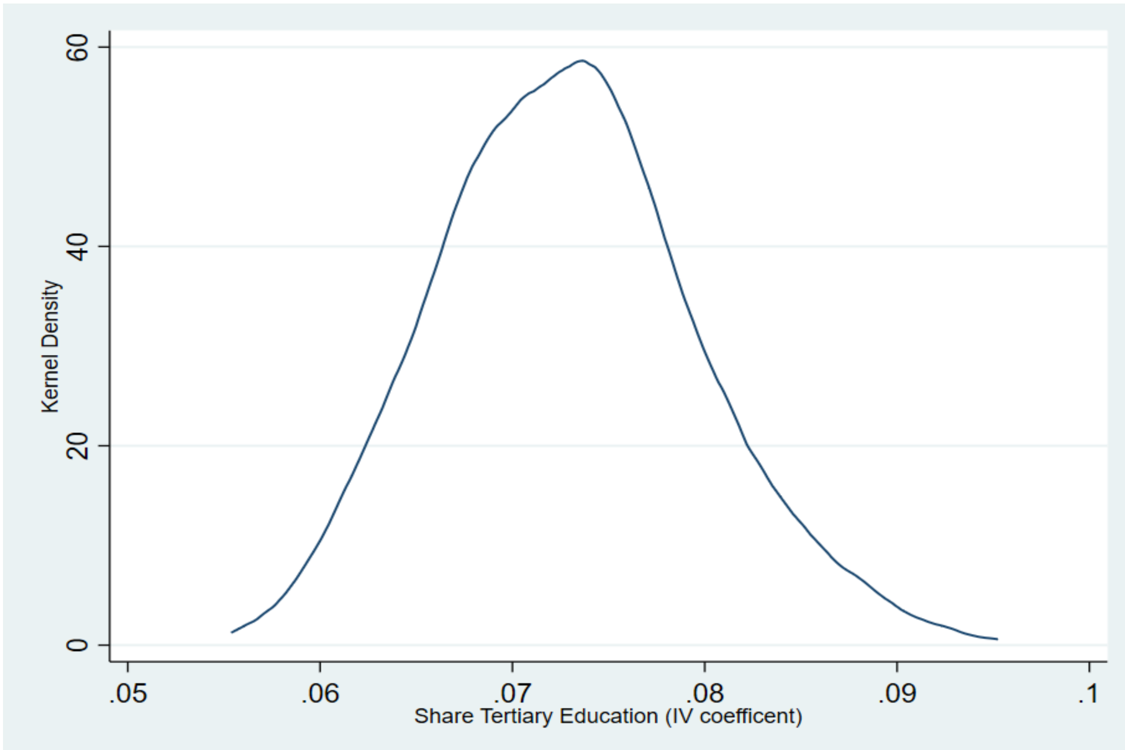
Notes: - The above graph shows the share of tertiary education in rural India over the last three decades for two age groups: individuals aged 25 or older and individuals aged 25 to 34. Data from the years 1991, 2001, and 2011 were obtained from the Population Census of India, while data from the year 2018-19 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



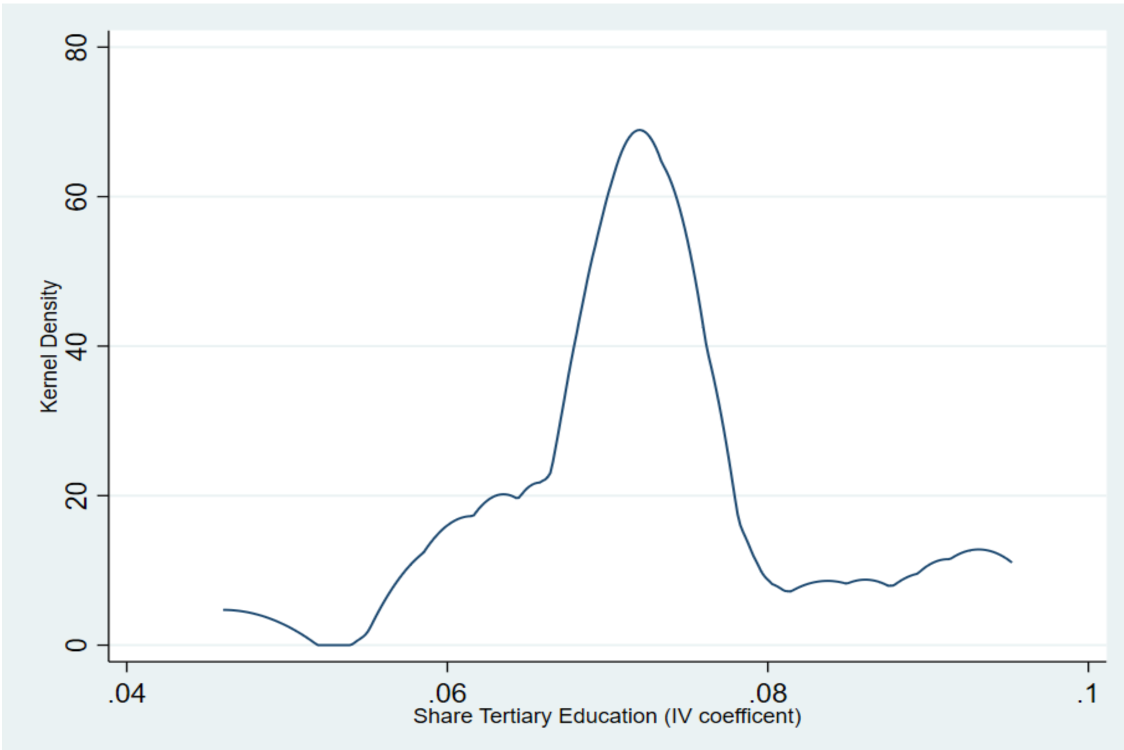
Notes: - Dark shading indicates a higher proportion of tertiary education in rural India, while light shading indicates a lower proportion.

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					
Nearest Distance from Catholic Missionary	457,967	71.33	46.61	0.156	420.2
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)
Dependent Variable: - Consumption per Capita (log)					
Share Tertiary & Above (pop 25+)	0.0148***	0.0126***	0.0108***	0.0113***	0.0112***
	(0.0007)	(0.0004)	(0.0004)	(0.0003)	(0.0003)
SC Proportion			-0.0721***	-0.0674***	-0.0706***
			(0.0053)	(0.0050)	(0.0049)
ST Proportion			-0.2758***	-0.2486***	-0.2466***
			(0.0073)	(0.0075)	(0.0076)
Population (log)			-0.0097***	-0.0154***	-0.0152***
			(0.0012)	(0.0013)	(0.0013)
Latitude				-0.1156***	-0.1076***
				(0.0078)	(0.0077)
Latitude Square				0.0028***	0.0026***
				(0.0002)	(0.0002)
Longitude				-0.2116***	-0.1755***
				(0.0289)	(0.0294)
Longitude Square				0.0012***	0.0010***
				(0.0002)	(0.0002)
Coastal (dummy)				-0.0097	-0.0185**
				(0.0072)	(0.0076)
Average River Length				-0.0004	-0.0005
				(0.0006)	(0.0006)
Average Height District				-0.0000	-0.0000
				(0.0000)	(0.0000)
Mean Rainfall (annual)				0.0002**	0.0003***
				(0.0001)	(0.0001)
Soil Quality (sandy)				-0.0035	-0.0016
				(0.0158)	(0.0154)
Nearest Distance from Big City				-0.0003***	-0.0002***
				(0.0000)	(0.0000)
Mean Temperature (annual)				0.0092***	0.0090***
				(0.0024)	(0.0024)
Village Area Square km (log)				0.0159***	0.0155***
				(0.0025)	(0.0024)
Fraction of Brahman (1931)					0.3428***
					(0.0919)
Fraction of Tribal (1931)					0.0720***
					(0.0269)
Fraction of Urban (1931)					0.2422***
					(0.0385)
Princely State (dummy)					0.0213***
					(0.0053)
Rail (1909)					0.0180***
					(0.0057)
Observations	427,242	427,242	427,242	427,242	427,242
R-squared	0.092	0.375	0.433	0.456	0.459
State FE	NO	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)
Panel (a): First stage - Share Tertiary & Above (pop 25+)					
Mean Subdistrict Distance Catholic Missionary (log)	-1.0136***	-0.4531***	-0.4669***	-0.1956**	-0.2263**
	(0.0864)	(0.0832)	(0.0819)	(0.0857)	(0.0880)
Kleibergen-Paap rk Wald F Statistic	137.692	29.684	32.502	5.206	6.612
Panel (b): Second stage - Consumption per Capita (log)					
Share Tertiary & Above (pop 25+)	0.0765***	0.0516***	0.0576***	0.0940**	0.0723**
	(0.0074)	(0.0113)	(0.0116)	(0.0406)	(0.0287)
Anderson-Rubin (AR) Test Statistic Chi2(1)	175.58***	35.08***	63.91***	27.97***	21.03***
Weak IV Robust 95% Confidence Interval	[0.0639, 0.0931]	[0.0342, 0.0825]	[0.0397, 0.0901]	[0.0506, ∞]	[0.0371, ∞]
Observations	427,242	427,242	427,242	427,242	427,242
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 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				
	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)
Panel (a): First stage - Share Tertiary & Above (pop 25+)									
Mean Subdistrict Distance Catholic Missionary (log)	-0.1486*	-0.2688***	-0.2445***	-0.2050*	-0.2365***	-0.2277***	-0.2199**	-0.2265***	-0.2262***
	(0.0888)	(0.0894)	(0.0894)	(0.1079)	(0.0869)	(0.0878)	(0.0869)	(0.0879)	(0.0878)
Kleibergen-Paap rk Wald F Statistic	2.799	9.031	7.470	3.610	7.411	6.723	6.407	6.650	6.645
Panel (b): Second stage - Consumption per Capita (log)									
Share Tertiary & Above (pop 25+)	0.0969*	0.0517***	0.0674***	0.0929**	0.0667***	0.0720**	0.0738**	0.0723**	0.0723**
	(0.0582)	(0.0190)	(0.0254)	(0.0466)	(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	School Index	Primary Health Center	Maternity Center	Post Office	Drainage System
Anderson-Rubin (AR) Test Statistic Chi2(1)	17.18***	14.65***	21.06***	21.67***	19.90***	21.10***	20.54***	21.01***	21.09***
Weak IV Robust 95% Confidence Interval	[0.0393, ∞]	[0.0254, ∞]	[0.0362, ∞]	[0.0468, ∞]	[0.0354, ∞]	[0.0372, ∞]	[0.0373, ∞]	[0.0371, ∞]	[0.0371, ∞]
Observations	403,697	415,762	417,659	282,677	424,664	427,242	427,242	427,242	427,242
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: - 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**	-0.2770***	-0.2402***
	(0.0880)	(0.0902)	(0.0911)	(0.0916)
Kleibergen-Paap rk Wald F Statistic	4.823	4.389	9.253	6.872
Panel (b): Second stage - Consumption per Capita (log)				
Share Tertiary & Above (pop 25+)	0.0649**	0.0615**	0.0608***	0.0681**
	(0.0308)	(0.0307)	(0.0210)	(0.0270)
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***	20.37***	19.39***
Weak IV Robust 95% Confidence Interval	[0.0270, ∞]	[0.0239, ∞]	[0.0333, 0.0952]	[0.0349, ∞]
Observations	427,242	427,242	393,658	401,130
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)
Panel (a): First stage - Tertiary Educated Member in Household (dummy)			
Mean District Distance Catholic Missionary (log)	-0.0145**	-0.0148**	-0.0148**
	(0.0064)	(0.0064)	(0.0064)
Kleibergen-Paap rk Wald F Statistic	5.073	5.382	5.369
Panel (b): Second stage			
	Agriculture Output per Unit of Land	Access to Technical Advice	Crop Diversification
Tertiary Educated Member in Household (dummy)	5.5924*	3.7984*	2.1627
	(3.3422)	(2.1885)	(1.4471)
Anderson-Rubin (AR) Test Statistic Chi2(1)	4.18**	5.24**	4.48**
Weak IV Robust 95% Confidence Interval	[0.4313, ∞]	[0.5922, ∞]	[0.1573, ∞]
Observations	29,457	29,596	29,652
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. Household Controls: Mean Age, SC (dummy), ST (dummy), Hindu (dummy), Household Size, # Male, # children ($\leq$ age 6), 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: Mechanism, IV Estimates - Employment Channel

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): First stage - Tertiary Educated (dummy)						
Mean District Distance Catholic Missionary (log)	-0.0077***	-0.0077***	-0.0077***	-0.0077***	-0.0077***	-0.0077***
	(0.0030)	(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	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	Skilled Agriculture
Tertiary Educated (dummy)	0.6166**	-0.2431	0.3735	0.2836	0.2172*	3.1550**
	(0.2957)	(0.3100)	(0.3214)	(0.2742)	(0.1214)	(1.5552)
Anderson-Rubin (AR) Test Statistic Chi2(1)	6.14**	0.90	1.04	1.07	4.54**	10.78***
Weak IV Robust 95% Confidence Interval	[0.1600, ∞]	[$-\infty$, 0.2111]	[-0.7336, 1.2516]	[-0.3569, ∞]	[0.0201, ∞]	[1.1229, ∞]
Observations	125,023	125,023	125,023	125,023	125,023	125,023
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. 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).

Table 14: 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	-0.0749	-0.4945***
	(0.1039)	(0.1411)	(0.1569)	(0.0882)
Kleibergen-Paap rk Wald F statistic	9.406	0.505	0.228	31.410
Panel (b): Second stage - Consumption per Capita (log)				
Share Tertiary & Above (pop 25+)	0.0867***	-0.0049	-0.2311	0.0803***
	(0.0292)	(0.0570)	(0.5096)	(0.0146)
Weak IV Robust 95% Confidence Interval	[0.04845, ∞]	[-∞, ∞]	[-$\infty, -0.0495] \cup$ [$0.1119, \infty$]	[0.0589, 0.1213]
Anderson-Rubin (AR) test statistic Chi2(2)	30.29***	0.01	10.31***	81.29***
Heterogeneity	Good Soil Quality	Bad Soil Quality	Bad Rainfall	Good Rainfall
Observations	216,303	210,939	214,845	212,397
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).

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.

Asadullah, M. N. and Rahman, S. (2009). Farm productivity and efficiency in rural bangladesh: the role of education revisited. *Applied economics*, 41(1):17–33.

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. World Bank Economic Review.

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

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

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.

Bezabih, M. and Sarr, M. (2012). Risk preferences and environmental uncertainty: Implications for crop diversification decisions in ethiopia. *Environmental and Resource Economics*, 53:483–505.

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.

Bozzola, M., Massetti, E., Mendelsohn, R., and Capitanio, F. (2018). A ricardian analysis of the impact of climate change on italian agriculture. *European Review of Agricultural Economics*, 45(1):57–79.

Bozzola, M. and Smale, M. (2020). The welfare effects of crop biodiversity as an adaptation to climate shocks in kenya. *World Development*, 135:105065.

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. *The 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.

Craig, B. J., Pardey, P. G., and Roseboom, J. (1997). International productivity patterns: accounting for input quality, infrastructure, and research. *American Journal of Agricultural Economics*, 79(4):1064–1076.

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.

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

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., Lagakos, D., and Waugh, M. E. (2014). Agricultural productivity differences across countries. *American Economic Review*, 104(5):165–170.

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. and Schoellman, T. (2018). Wages, human capital, and barriers to structural transformation. *American Economic Journal: Macroeconomics*, 10(2):1–23.

Hicks, J. H., Kleemans, M., Li, N. Y., and Miguel, E. (2017). Reevaluating agricultural productivity gaps with longitudinal microdata. Technical report, National Bureau of Economic Research.

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 Valencia Caicedo, F. (2017). Engineering growth: innovative capacity and development in the americas.

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

McMillan, M., Rodrik, D., and Sepulveda, C. (2017). Structural change, fundamentals and growth: A framework and case studies. Technical report, National Bureau of Economic Research.

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.

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

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

Porzio, T. and Santangelo, G. (2017). Structural change and the supply of agricultural workers.

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.

Roca, J. D. L. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

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

Schultz, T. W. (1964). Changing relevance of agricultural economics. *Journal of Farm Economics*, 46(5):1004–1014.

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. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

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.

Taylor, J. E. and Martin, P. L. (2001). Human capital: Migration and rural population change. *Handbook of agricultural economics*, 1:457–511.

Vollrath, D. (2007). Land distribution and international agricultural productivity. *American Journal of Agricultural Economics*, 89(1):202–216.

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.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785.

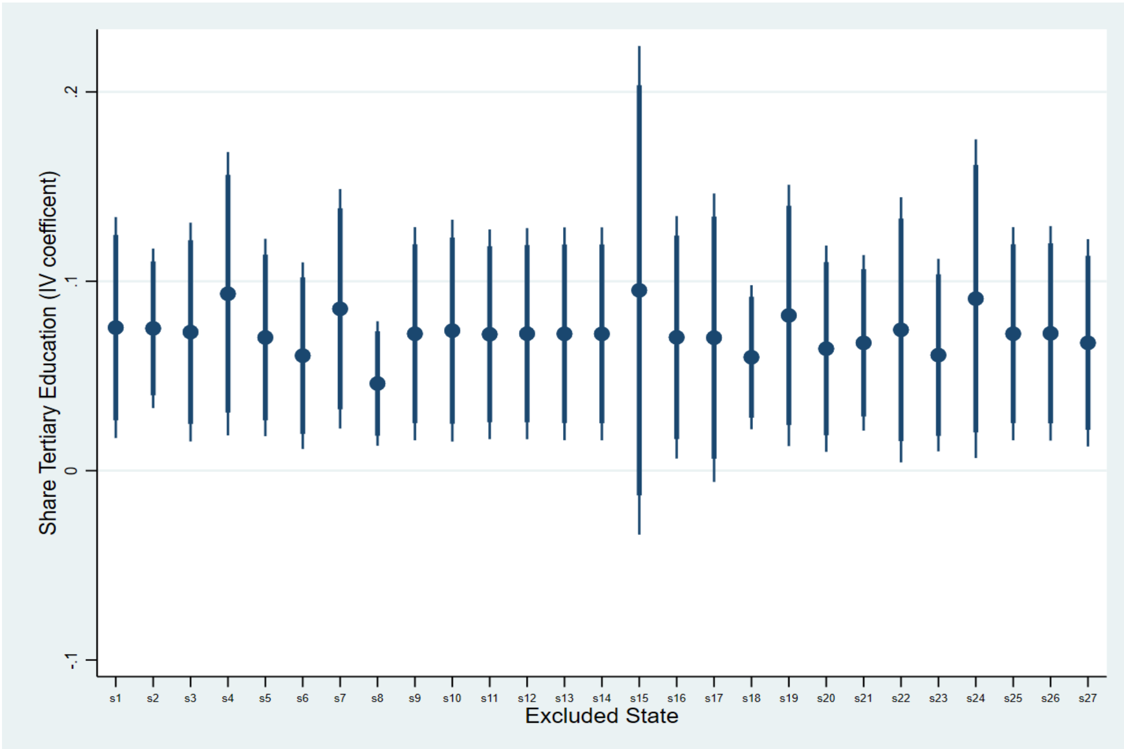
Zaveri, E., Russ, J., and Damania, R. (2020). Rainfall anomalies are a significant driver of cropland expansion. *Proceedings of the National Academy of Sciences*, 117(19):10225–10233.

Appendices

A 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.

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, Including Lower Level of Education

	(1)	(2)	(3)	(4)
Panel (a): First stage - Share Tertiary & Above (pop 25+)				
Mean Subdistrict Distance Catholic Missionary (log)			-0.2263**	-0.2082**
			(0.0880)	(0.0833)
Share Below Graduate (pop 05+)				0.0497***
				(0.0022)
Kleibergen-Paap rk Wald F statistic			6.613	6.247
Panel (b): Second stage				
	OLS	OLS	IV	IV
Share Tertiary & Above (pop 25+)	0.0112***	0.0099***	0.0723**	0.0717**
	(0.0003)	(0.0003)	(0.0287)	(0.0311)
Share Below Graduate (pop 05+)		0.0034***		0.0004
		(0.0001)		(0.0016)
Anderson-Rubin (AR) Test Statistic Chi2(1)			21.01***	21.01***
Weak IV Robust 95% Confidence Interval			[0.0370, ∞]	[0.0360, ∞]
Observations	427,224	427,224	427,224	427,224
Controls				
R-squared	0.460	0.491		
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 A2: OLS Results, Catholic Missionary on Alternative Channels

	Under 5 Mortality Rate	Share Irrigated Land	Violent Crimes per Capita	Confidence Panchayats	School Index	Primary Health Center	Maternity Center	Post Office	Drainage System
Mean Subdistrict Distance Catholic Missionary (log)	-0.4727	-0.6599	-0.0000	-0.0028	-0.0033	-0.0001	-0.0060	-0.0009	-0.0012
	(0.6124)	(0.4020)	(0.0000)	(0.0042)	(0.0033)	(0.0008)	(0.0042)	(0.0013)	(0.0028)
Observations	430,816	444,855	444,268	302,716	449,872	456,529	456,529	456,529	456,529
R-squared	0.782	0.570	0.404	0.486	0.273	0.070	0.089	0.098	0.241
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: - 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 A3: 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 kilometers Source: https://eogdata.mines.edu/products/vnl/ and Map of India, processed by Arc GIS
Main 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
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 of India 2011
Nearest Distance from Catholic Missionary	
Fraction of Brahman (1931)	Share of Brahman population 1931 Source: Census 1931
Fraction of Tribal (1931)	Share of tribal population 1931 Source: Census 1931

Fraction of Urban (1931)	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
Fraction of Urban (1901)	Share of urban population in 1901 Source: Census 1901
Fraction of Brahman (1901)	Share of Brahman population in 1901 Source: Census 1901
Fraction of Lower Castes (1901)	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)	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 Source: District Gazetteers of India
School (1901)	Number of schools in 1901 Source: District Gazetteers of India

Income Tax per Capita (1901)	Income tax revenues per capita in 1901 Source: District Gazetteers of India
Under 5 Mortality Rate	Under 5 Mortality Rates of a district in 2001 Source: Census, 2001