UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2023/89

Transforming rural economies through tertiary education

Evidence from India

Amparo Castelló-Climent,[1] Abhiroop Mukhopadhyay,[2] and Ravinder[2]

July 2023

Abstract: This paper analyses the role of tertiary education on rural development. Using census data on villages in India for 2011, we find that skilled workers have had an important impact on rural prosperity. A 1 percentage point rise in the share of the village population with tertiary education raises per capita consumption by around 7.2 per cent. Our results are robust to alternative measures of income and are not confounded by better institutions. Among the mechanisms at work, we find that households with tertiary-educated members register higher agricultural gross income per unit of land. Their contribution to agriculture is also suggested by satellite-based spatial net primary productivity measures. We inform the larger literature on structural transformations and contend that such transformations can also happen within rural areas.

Key words: rural, transformations, tertiary education

JEL classification: I25, J24, O11, O13

Acknowledgements: We would like to thank the conference attendees of the 17th ACEGD 2022 (ISI Delhi), the 27th ACCDE (Jadavpur University), the Conference on 'Socio-Economic Inequality in India' (CSH-New Delhi), the Association for the Economics of Education conference, and AGW 2022 (APU Bangalore), and participants at the CAGE working paper seminar (Warwick University) and IIT (Jodhpur). We would also like to express our appreciation for the financial support received from the Spanish Ministry of Economy and Competitiveness under the PGC2018-101161 project and the Planning and Policy Research Unit-ISI (Delhi). This work was done while Ravinder was a visiting PhD fellow at UNU-WIDER in Helsinki, and he is thankful to Professor Kunal Sen for his guidance and mentorship during this period.

[1] University of Valencia, Valencia, Spain; Indian Statistical Institute, Delhi, India; corresponding author: ravinder20r@isid.ac.in

This study has been prepared within the UNU-WIDER project Academic excellence.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-397-0

https://doi.org/10.35188/UNU-WIDER/2023/397-0

Typescript prepared by Gary Smith.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

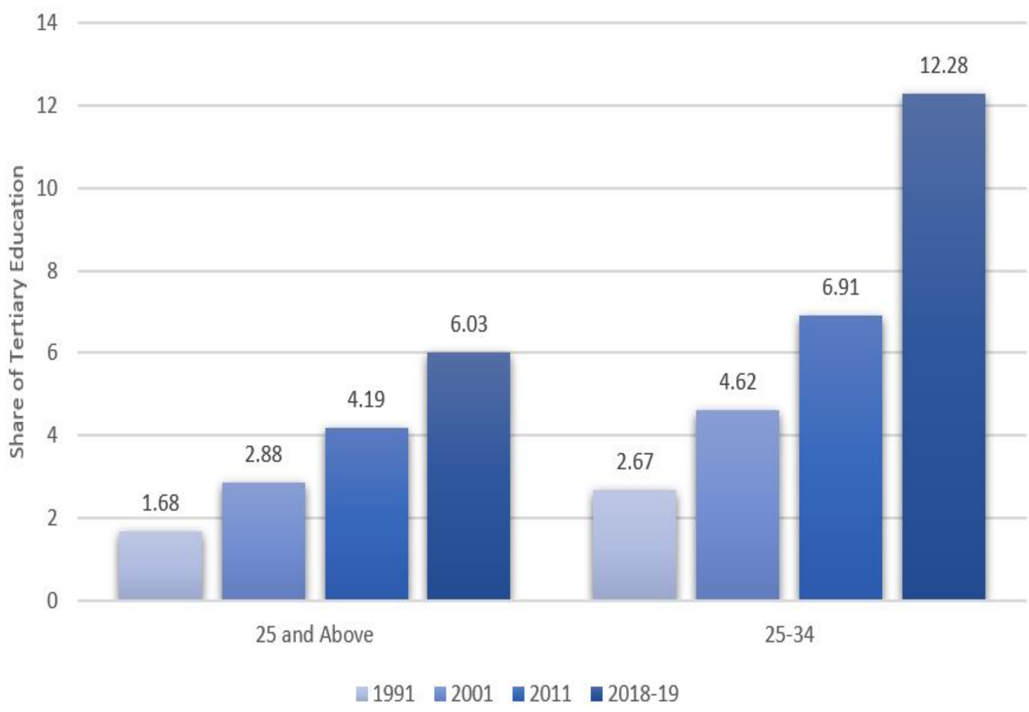
The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Economic development is associated with a process of structural transformation where the demand for workers outside the agriculture sector increases. The movement of labour 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 movement comes with a transfer of labour from rural to urban areas (Gennaioli et al. 2013; Harris and Todaro 1970; Lewis 1954). In the less developed countries, however, despite the increased levels of education in recent decades, most of the population still resides in rural areas. This is particularly relevant in the case of India where, according to the 2011 census, approximately one-third of tertiary-educated individuals reside in rural areas. Figure 1 shows a tremendous increase in the number of tertiary-educated individuals living in rural areas in recent decades. For the age group of 25–34 the share of tertiary-educated people residing in rural areas increased from 6.91 per cent in 2011 to 12.29 per cent in 2018. In this paper we analyse the role of tertiary-educated people in rural development and show that skilled workers have had an important impact on rural prosperity. We find that a 1 percentage point increase in the share of village population with tertiary education raises per capita consumption by around 7.2 per cent.

Figure 1: Share of tertiary education in rural India, 1991–2019



Note: this 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 25–34. Data for the years 1991, 2001, and 2011 were obtained from the Population Census of India, while data for the year 2018–19 were obtained from the Periodic Labor Force Survey (PLFS), which is a nationally representative survey.

Source: authors' illustration based on census data and the PLFS.

We estimate the impact of a higher share of tertiary-educated individuals in the village labour 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] One concern, however, is the lack of available data on income per capita. To overcome this limitation we rely on data on consumption, as measured by village-level mean per capita consumption from the Socioeconomic High-resolution Rural–Urban Geographic Data Platform for India (SHRUG), 2011. This measure is imputed consumption rather than actual consumption. To show our results are robust we use 1,000 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, using the Visible Infrared Imaging Radiometer Suite (VIIRS) Night Lights (2015), extracted at the village level.

Human capital is endogenous and evolves hand in hand with income levels, and with 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 a 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 the state governments, making it unlikely that institutional differences drive our results. However, since the skills of the decision-makers at the local level could affect development (Casey et al. 2023), we check the robustness of the result by controlling for the provision of public goods at the village level. To account for other potential omitted variables bias, we control for a broad set of current demographic controls, geographical characteristics, and historical variables.

We address any remaining threats to exogeneity of tertiary education by using an instrumental variable. We identify the impact of tertiary education by using the variation in the historic location of Catholic missions 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. While this change was achieved through an expansion of Catholic educational institutions following India's independence, they were located around the original location of the Christian mission.

An extensive reading of Catholic history suggests that Catholicism in India was unregulated by the Vatican in the period from the sixteenth century to the 1930s. Hence, which district Catholic missions located in was dependent on a myriad of idiosyncratic strategies by individual missionaries. Some factors do predict the location of Catholic missions—whether the district was on the coast, whether the district had a railway line passing through, and whether the district was tribal. Most importantly, locations were not determined by incomes in the areas, as predicted by income tax collections (Castelló-Climent et al. 2018). Hence, such locations, conditional on controlling for the predictors, are likely to be exogenous. We extend this idea to measure the locations of missions as of 1911, relative to the locations of the villages of India.

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—we therefore use the mean distance to the nearest Catholic missionary, averaged over all villages in a sub-district, as our instrument (for 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 for the distance of a village to the nearest city and the historical urbanization rate in all our regressions. We also show in robustness checks that our results remain

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

[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.

unchanged when we account for the sub-district urbanization rate or account for urban wages in the district.

The mean distance from a Catholic mission is highly correlated with the rural share of tertiary completion—the closer the average village in a sub-district to the historical location of the Catholic mission, the higher the tertiary completion rate. Using this exogenous source of variation for the current share of tertiary-educated people, we find a positive and statistically significant effect of tertiary education on village economic development, proxied by mean consumption per capita. A 1 percentage point increase in the share of the population with tertiary education increases the log mean consumption per capita by 7.2 per cent.

We provide evidence that our instrument meets the exclusion restriction by a host of sensitivity checks. We show that the mean distance from Catholic missions is uncorrelated with many markers of development at the district level (under-five mortality rates, share of land that is irrigated, confidence in local village councils, the number of violent crimes per capita) and at the village level (provision of schools, primary health centre, maternity centre, post office, availability of drainage). Moreover, our results do not change even if we control for these variables.

We then analyse the mechanisms that help explain the way tertiary education positively affects village prosperity. Given that most of the tertiary-educated work in rural areas, and rural economies are largely based on agriculture, we analyse whether tertiary-educated workers have had an impact on agricultural productivity. There are no village-level data on crop yields. Instead, we use micro-data on household agricultural profits from the nationally representative survey of agricultural households (the Situation Assessment Survey of Agricultural Households, collected by the National Sample Survey Organization (NSSO) in 2013), as well as satellite images. We look at the impact of having a tertiary-educated member in the household on two measures of agricultural income—the value of agriculture gross value of crops per unit of land cultivated and crop diversification. We find that households with tertiary-educated members show higher agricultural gross value per unit of land. Additionally, they are more likely to diversify their crops (Birthal et al. 2015; Hazra 2001; Joshi et al. 2004) and have better access to technical knowledge related to agriculture. These findings suggest that tertiary-educated individuals play a positive role in adopting new and more productive technologies. In addition, 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-educated areas leads to higher agricultural productivity. The evidence is in line with previous findings that show that in the Indian context the human capital of labour is an important factor facilitating the adoption of new and more productive technologies (Foster and Rosenzweig 1995, 1996). However, in contrast to previous studies that point out the importance of primary education during the era of the Green Revolution as the background, we use a more recent context where agriculture is less traditional and rural tertiary completion rates are five times higher. Our results are in line with those of Schultz (1975), which point to the importance of human capital, especially when agriculture is non-traditional. Interestingly, some of our effects also emanate from having tertiary-educated members in the household, even though they may not work in agriculture—it is plausible that they are sources of information for new technology and markets.

Ideally we would decompose village income into its constituent elements to see what drives the impact of tertiary education. Though we provide some evidence above on the impact on agricultural income, we are unable to do a full decomposition due to lack of a representative data set at the village level. Hence, we provide some suggestive evidence by showing the impact of tertiary education on the probability of being in skilled occupations—providing jobs that are productive and pay well.

To show this, we rely on micro-data from the Employment–Unemployment Survey conducted by the NSSO in 2009. Results from rural India indicate that, as compared to less educated workers, tertiary-educated workers have a greater probability of 'regular wage' jobs in the private sector. In contrast to the public sector, private wages reflect the value of the marginal product—so we contend that not only does tertiary education lead to higher income, but these are accompanied by higher productivity. Further, tertiary education causes individuals to become skilled agriculturalists—we capture part of this effect in the results on agricultural income.

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, we show that the tertiary-educated in the labour force are more likely to have such jobs than the less educated—this suggests a model of structural transformation that does not rely on moving educated people through migration to urban areas. Instead, this relies on educated people living in the village but commuting to nearby towns. However, this is not the only source of village prosperity since the proportion of tertiary-educated who commute daily to the nearest urban centre is only 9 per cent. Together with the results on agricultural income, this suggests that a significant part of the prosperity is generated within the village.

Our paper is related to the literature that points to highly educated individuals as drivers of development and growth. Historical evidence shows that skilled workers played a key role in the Industrial Revolution (Maloney and Valencia Caicedo 2017; Meisenzahl and Mokyr 2011; Squicciarini and Voigtländer 2015). Using contemporary data, Gennaioli et al. (2013) find that managerial education plays a bigger role in explaining regional development than workers' human capital. The emphasis on tertiary education has been particularly important in India. High-quality engineering and technology-oriented institutions of higher education have been the aim of all Indian governments over the years. As a result, the share of population with tertiary education is higher in India than in China. In 2015 the share of the working-age population (15–64 years old) with tertiary education was 12.68 per cent in India compared to 7.27 per cent 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 et al. 2018; Castelló-Climent and Mukhopadhyay 2013). 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 Indian villages.

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). 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).[3] Higher education leads individuals to move from the agriculture sector to manufacturing and services. Typically, such movement comes with the migration of labour from rural to urban areas (Gennaioli et al. 2013; Harris and Todaro 1970; Lewis 1954). 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 specialized jobs are concentrated in cities. In this paper we show evidence of a different structural transformation driven by skilled workers, one that happens within rural areas and within the agriculture sector.

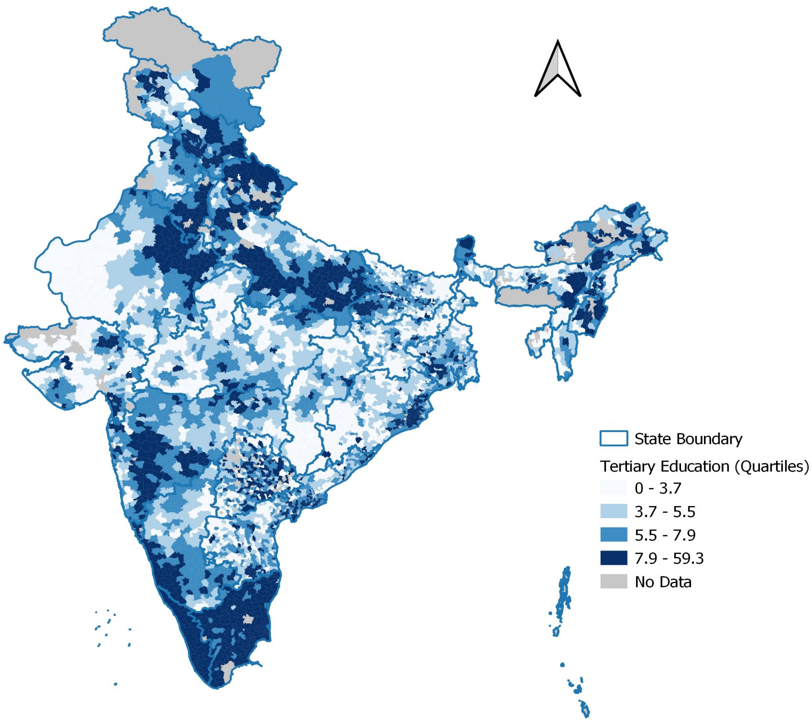
The structure of the paper is as follows. In Section 2 we describe the data and set out the empirical methodology. In Section 3 we present the main results. We describe several robustness checks in Section 4. We analyse the mechanisms through which tertiary education influences rural prosperity in Section 5. Our conclusions are summarized in Section 6.

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

2 Data and empirical strategy

This paper uses data on the census of villages in India. India had around 650,000 villages in 2011. The analysis amalgamates data on variables from different data sets that need to be merged, often based on village names, which are often recorded slightly differently across data sets. Since perfect matching is not possible for some villages, the final data set covers around 450,000 villages across 27 states.[4] The analysis focuses on the share of the population that has completed tertiary education—that is, those with university-level degree/diploma and above. We source these data from the Socio-Economic and Caste Census (SECC), collected in 2011.[5] Precisely, we define *Share_ter* as the share of population who have completed tertiary education. To ensure that our measure is not right censored by children who are not old enough to reach their highest education level, we use the population aged 25 and above as the relevant population.[6] Figure 2 shows the spatial distribution of the share of tertiary education across sub-districts in rural India. Lighter shades represent a lower share of tertiary education. The figure shows that southern and some northern Indian states have a larger share of people with tertiary education and that there is significant heterogeneity within states.

Figure 2: Share of tertiary education at the rural India sub-district level



Note: dark shading indicates a higher proportion of tertiary education in rural India, while light shading indicates a lower proportion.

Source: authors' calculations based on data from the SECC.

[4] 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 which are perfectly matched.

[5] 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.

[6] 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 age 22, the measurement error in this variable, though present, is unlikely to be large.

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 non-availability of income per capita data at the village level. To overcome this limitation we use different measures of development. Our primary measure of development is a consumption measure, estimated by Asher et al. (2021). This proxy is available as part of the SHRUG data library for village-level data. This measure is generated using small-area estimates following the methodology in Elbers et al. (2003).[7] As is standard in the literature, we use the log of mean consumption per capita as our dependent variable.[8]

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 per cent.[9] 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 controls in our model, and estimate the following specification:

$$\ln\_Cons\_per\_Capita_{vds} = \alpha_s + \theta Share\_Ter_{vds} + \rho' C_{vds} + \pi' G_{vds} + \delta' H_{ds} + \varepsilon \tag{1}$$

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; namely, 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 (more on this when we discuss our instrument).[10]

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 the 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; we use the latitude and longitude of the centroid of the village and their squares,[11] the average length of rivers that pass through the district (kilometres) and the average height (elevation) of the district (kilometres),[12]

---

[7] 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 Asher and Novosad 2020: Table A6) at the household level, contained in both data sets. The SECC 2011 household data are only available to the SHRUG team; the public data reports these measures for the village level, which we use in our paper.

[8] Since consumption is imputed, it has a distribution. For our analysis we use the mean of the imputed value. However, 1,000 draws of village-level per capita consumption are also reported in SHRUG, which we use in the robustness check (Section 4).

[9] The share of tertiary-educated may suffer from some measurement error as the maximum value is 100 per cent. The mean for the top 1 per cent is 27.2 per cent. We do not delete these high values as relatively little 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 per cent.

[10] In India, a sub-district, 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 sub-district usually consists of several villages or towns grouped together for administrative purposes.

[11] Extracted from the village shape files available from Maps of India.

[12] Both sourced from http://www.diva-gis.org and extracted using Arc GIS tools.

annual rainfall and temperature in a village,[13] the area of the village,[14] an indicator variable for districts with any coastal boundary which is based on district maps from the Census of India 2011, the proportion of soil that is sandy in a district,[15] and the distance to the nearest million-plus city.[16]

To capture differences in initial conditions across districts, which 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 the Census of India 1931 and includes the share of the district population that is urban, the share of the 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.[17] These variables account for the evolution of other contemporaneous variables. We report summary statistics for the main variables 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, better property rights, etc., which also impact the economy. It is therefore hard to 'identify' the causal impact of education alone. To address this challenge we analyse the impact of higher education by using the variation in the historic location of Catholic missions in India circa 1911. As explained by Castelló-Climent et al. (2018), Catholic missions changed the supply of and preference for higher education due to their emphasis on providing quality education. While 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 flock to read the Bible—hence they promoted literacy and lower levels of education (Mantovanelli 2014).[18]

We source information on the location of Catholic missionaries from a map published in the first edition of *Atlas Hierarchicus*, which marks the name of every place in India where there was a Catholic mission or missionary in 1911. Catholicism was not a coordinated movement in India. While the early missionaries followed Portuguese conquest, missionaries also settled in the interior away from Portuguese strongholds. Our reading suggests the individual preferences of missionaries played a role (for more, see Castelló-Climent et al. 2018) over and above some predictable factors that we control for. 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 tribal population. Reassuringly, the location of Catholic missionaries was uncorrelated with the share of Brahmans and income tax revenues per capita, a proxy for income.

[13] Sourced from the Climatic Research Unit Database Version 3.22 (CRU), University of East Anglia Climatic Research Unit, available at http://catalogue.ceda.ac.uk/uuid/3f8944800cc48e1cbc29a5ee12d8542d.

[14] Sourced from the Census of India 2011

[15] Sourced from 'Soils of India', a database of the NBSS and LUP, the Indian Council of Agricultural Research

[16] There are 53 such cities according to the Census of India 2011; distances are calculated from each village using Arc GIS.

[17] 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 cross-walk between districts in 1931 and 2001 was available in previous data constructed by Castelló-Climent et al. (2018), and we extend that to 2011 using the Atlas of India from the Census of India 2011. The data on railways is sourced from a map in the *Indian Administrative Report on Railways*, which is further processed by Arc GIS.

[18] Protestant missionary location cannot, however, 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).

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 km^2 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 km^2 (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 2					
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

Source: authors' calculations.

We extend this idea to look at the location of Catholic missions relative to each village. This is done using Arc GIS using the centroid of the village area and the exact location of the Catholic missions. Further, we use an aggregate measure—the mean distance of a village from the nearest Catholic mission, averaged over all the villages in a sub-district. This removes any possible bias specific to any particular village location. For an easier reference, we refer to this instrument as the mean distance to Catholic mission and use its logarithmic value in our regressions.

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

A critical requirement for identification is that the instrument has to influence the measure of village development only through the share of the 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 Catholic missions located in 1911 was not entirely exogenous, some predictable variables explain their location. Our instrument varies at the sub-district level. In Table 2 we regress the distance of Catholic missions on a broad set of geographical and historical variables. Analogous to results for district location, the results indicate that distance to Catholic missions is lower 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 these urban centres. Hence, distances to cities and towns are positively correlated to the mean distance to Catholic missions. Further, when we control for 1901 Census variables we also find that such distances are smaller when the district was more urbanized. This is partly mechanical since the larger the urban sprawl or the more urban centres there are, the closer are the rural areas. While 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 reflects also the opposite selection, giving further evidence to the idiosyncratic strategies of the Catholic missionaries. Reassuringly, income tax per capita (a good proxy for income) has no correlation with our instrument, nor does the proportion of Brahmans 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, mean distance to Catholic missions in 1911 is exogenous and can be used to examine the impact of tertiary education on village development. In the next section we provide evidence showing that the closer the average village to a Catholic mission, the higher the share of the population with higher education in 2011. In Section 4 we provide further evidence on the validity of our instrument.

Table 2: OLS results, determinants of Catholic missionary location

	(1)	(2)	(3)
Dependent variable: Mean sub-district 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 km^2 (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

Note: robust standard errors are in parentheses and clustered at the sub-district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: authors' calculations.

3 Main results

We start off with a descriptive OLS, regressing the log of village-level per capita consumption on the share of the adult population with tertiary education. The results are reported in Table 3, 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. The finding, however, 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. However, when we control for the share of the population with education below the graduate level our results remain unchanged in Table B1.[19]

Thus there is a strong correlation between the proportion of adult individuals with a college education and greater economic prosperity in Indian villages.

In order to make causal inferences, we move next to our IV results. Table 4 displays the results of the IV estimates using the log of mean distance to Catholic missions as the source of exogenous variation. Panel (a) displays the first-stage results. The most parsimonious specification is in column 1. The coefficient of Catholic mission is negative and statistically significant at the 1 per cent level. The coefficient remains significant when we control for state fixed effect and current, geographical and historical variables in columns 2–5. The results indicate that villages that were located farther away from a Catholic mission in 1911 have a lower share of the population with tertiary education in 2011. While 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 instruments as well as report weak IV robust 95 per cent confidence intervals 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 (columns 1–5). The coefficient with the full set of controls (column 5) implies that a 1 percentage point increase in the share of tertiary-educated individuals increases per capita consumption by 7.2 per cent.

As noted in our OLS results, it could be argued that the share of tertiary education is picking up the impact of other levels of schooling. In Table B1 we address this issue by including the share of lower education. As described earlier, this variable includes the share of people who have some level of education but do not have a tertiary 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, which strongly suggests that omitting lower levels of education is not driving our results on tertiary education.

Hence tertiary education completion has a large impact on village prosperity.[20] Before we move on to what may be the mechanisms at play, we provide some evidence that our results are robust.

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

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

As mentioned earlier, we do not 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).

[20] 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 claims on the impacts of lower levels of education.

Table 3: 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 km^2 (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

Note: robust standard errors are in parentheses and clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: authors' calculations.

Table 4: IV results, log consumption per capita

	(1)	(2)	(3)	(4)	(5)
Panel (a): First stage: share tertiary and above (pop 25+)					
Mean sub-district 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 and 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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

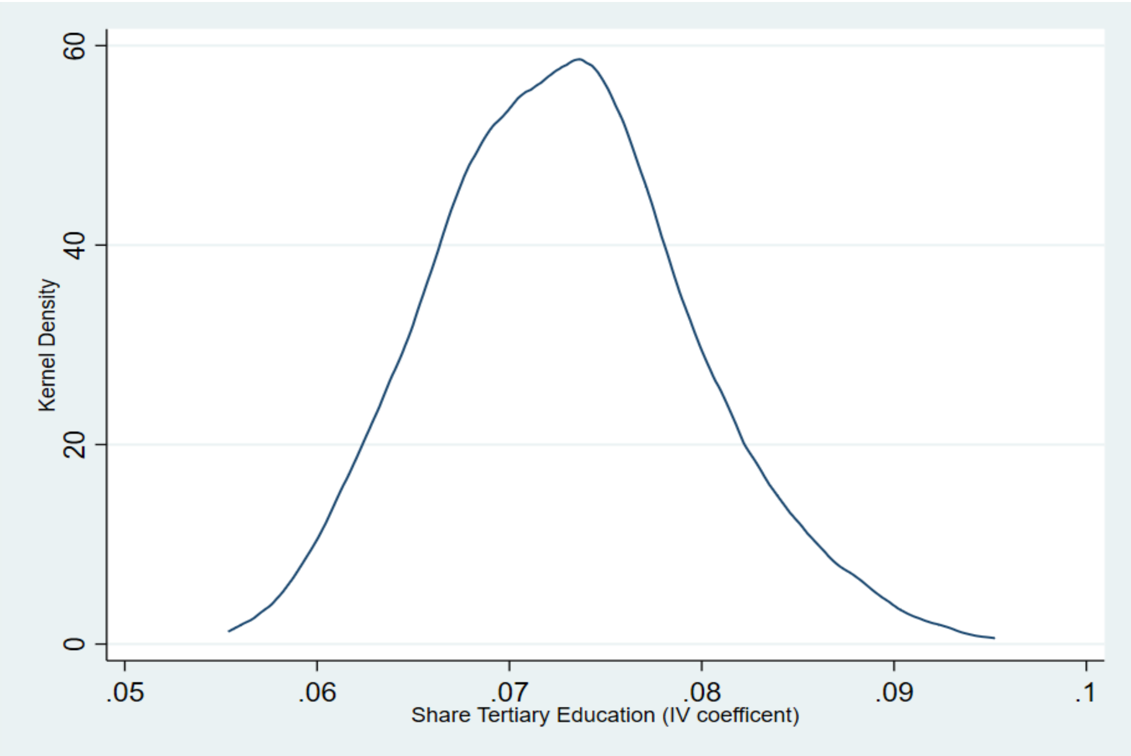
4 Robustness

In the previous section our IV estimates suggest a strong and significant causal relationship between tertiary education and village prosperity, as measured by log consumption per capita. In this section we address various concerns regarding our estimates and provide results of various sensitivity checks to bolster our claim.

4.1 Imputed consumption

The first concern is that our dependent variable is an imputed consumption measure rather than an actual consumption, and thus is susceptible to estimation error. To address this concern, a kernel density of the main IV specification was plotted (Silverman 1986), utilizing 1,000 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, inclusive of all controls in Table 4, column 5, was 0.0723, which is in close proximity 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.

Figure 3: Robustness, IV estimates: 1,000 replications of consumption per capita (log)



Note: the kernel density plot displays the distribution of the IV coefficients for the share of tertiary education, using 1,000 replications of bootstrapped log consumption per capita as the dependent variable.

Source: authors' calculations.

4.2 Alternative measure of income

Mean consumption per capita is a model-generated prediction and may be wrong if the prediction model is mis-specified. We check the robustness of the results with another proxy for income: nighttime light density, in line with the literature (Alesina et al. 2016; Henderson et al. 2012; Michalopoulos and Papaioannou 2013), as shown in Table 5. Panel (a) shows the results for the first-stage regression. The coefficient of log Catholic mission remains significant at the 1 per cent level, with the F statistic ranging around 6–8. Additional tests, such as the Anderson–Rubin test and the weak IV robust 95 per cent 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 VIIRS annual data for the year 2015.[21] Panel (b) shows the second-stage results. The coefficient of the share of tertiary education is positive and statistically significant at the 1 per cent level with all measures of income. Column 1 shows the results with the full sample. In the case of rural India, however, there is a problem with low light density at night, resulting in many zero values and some outliers as well.[22] To address the issue of extreme values we trim outliers in columns 2 and 3 by 1 per cent and 2.5 per cent (on both sides), respectively. The coefficients of the share of tertiary education remain positive and statistically significant. The economic effect when we use nighttime light data is very high and sensitive to trimming.

Table 5: Robustness, IV results: alternative measures of income

	(1)	(2)	(3)
Panel (a): First stage: share tertiary and above (pop 25+)			
Mean sub-district 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 and above (pop 25+)	0.1959***	0.1695***	0.1410***
	(0.0667)	(0.0614)	(0.0547)
Anderson–Rubin (AR) test statistic $\chi^2(1)$	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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

4.3 Skewed distribution

Another concern is related to the distribution of the dependent and the main independent variables (share of tertiary education). The distribution of the main independent variable is skewed and has a considerable number of zero values. Likewise, the VIIRS nighttime lights also exhibit a skewed distribution, with many zeros and some outliers. This exacerbates the violation of the assumption of normality for the

[21] The availability of VIIRS nighttime light annual data at the pixel level for 2015 provides the closest approximation to 2011. In contrast to an 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 the background non-light values have been set to zero (Gibson 2021; Gibson et al. 2021). It is important to note that aggregating monthly data may not accurately represent the data, as it does not take into account seasonal differences. The VIIRS nighttime lights are an improvement over DMSP-OLS (Wu and Wang 2019), which is used in Chanda and Cook (2022), although they have used monthly data.

[22] 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.

dependent variable. To address this issue we transform these variables using an inverse hyperbolic sine (IHS) function. It has been shown that in cases with many zero values or outliers the IHS transformation is preferred over the natural logarithm transformation (Bellemare and Wichman 2020). These authors also provide a formula for computing the elasticity in this type of transformation. The results, displayed in Table 6, demonstrate that the first-stage results are significant at the 5 per cent level across all columns in panel (a). In panel (b) the coefficient of both the share of tertiary education and its IHS transformation are positive and statistically significant, also at the 5 per cent level. 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 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 plotted 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 Appendix Figure B1. 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 one 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 1 per cent. This trimming process involves removing the top and bottom 1 per cent of values from these variables. After trimming we present the findings in Table 8. Both the first and second stages in our analysis remain significant. This indicates that our results are not driven by extreme values.

Table 6: Robustness, IV results: inverse hyperbolic sine transformation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): First stage								
	Share tertiary and above (pop 25+)	Share tertiary and above (pop 25+)	Share tertiary and above (pop 25+) (IHS)	Share tertiary and above (pop 25+) (IHS)	Share tertiary and above (pop 25+)	Share tertiary and above (pop 25+)	Share tertiary and above (pop 25+) (IHS)	Share tertiary and above (pop 25+) (IHS)
Mean sub-district 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 and above (pop 25+)	1,056.6137**	0.0723**			0.4290***	0.2500***		
	(431.9170)	(0.0287)			(0.1525)	(0.0852)		
Share tertiary and above (pop 25+) (IHS)			6,486.5565**	0.4439**			3.0006**	1.7486**
			(2,723.6265)	(0.1815)			(1.2753)	(0.7198)
Anderson–Rubin (AR) test statistic $\chi^2(1)$	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, ∞]	[3,143.44, ∞]	[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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

Table 7: Robustness, IV results: geographical selection

	(1)	(2)	(3)	(4)
Panel (a): First stage: share tertiary and above (pop 25+)				
Mean sub-district 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 and above (pop 25+)	0.0723**	0.0398***	0.0410***	0.0356**
	(0.0287)	(0.0152)	(0.0144)	(0.0154)
Anderson–Rubin (AR) test statistic $\chi^2(1)$	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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

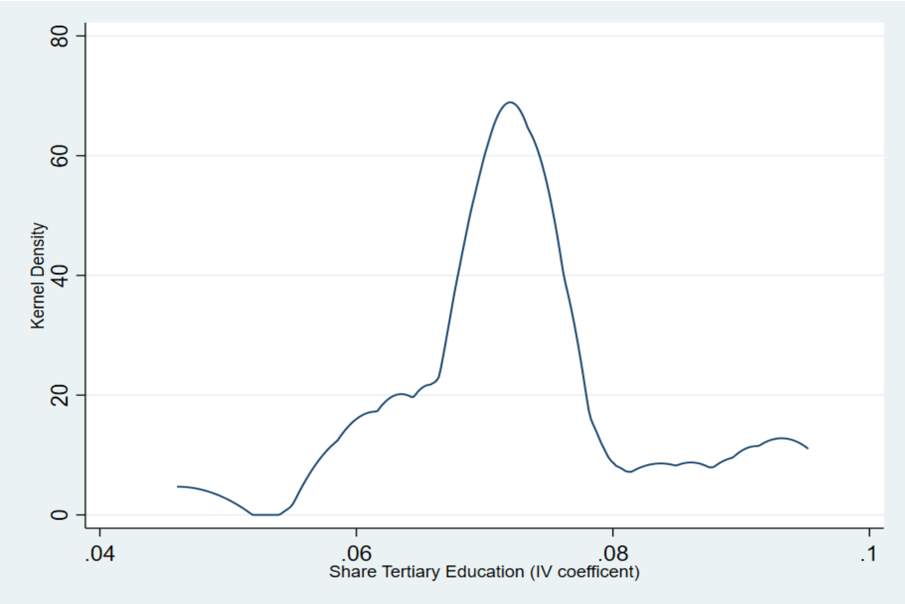
Table 8: Robustness, IV results: trimming

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel (a): First stage: Share tertiary and above (pop 25+)									
Mean sub-district 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 and 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 $\chi^2(1)$	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

Note: X: share tertiary and above (pop 25+) and Y: consumption per capita (log). Robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

Figure 4: Robustness, IV estimates: excluding one state



Note: 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.

Source: authors' calculations.

4.7 Exclusion restriction and other channels

In this analysis we show that our results are not confounded by other factors, proxies of institution, or other plausible missionary activity. The first set of potential confounders that we consider are health/medical interventions that often go hand in hand with those for education. To do so, we look at a reliable district-level indicator of health (under-five mortality rate) and supply of institutions at the village level (village-level access to primary health centre as well as maternity centre). Further, we look at some proxies for the quality of institutions in a district: 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 village-level amenities: 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.[23] To show exclusion, first, we show that our IV is not correlated with these potential confounders. Results in Table B2 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 Table 9. This suggests that our analysis is not confounded by these variables.

We have noted earlier that missionaries settled in urban centres—hence there may be concerns that our impacts are driven solely by urbanization. To rule out this channel, we control for proxies related to urbanization such as district urbanization rate, sub-district urbanization rate, district-level urban wages for the tertiary-educated (relative to those for the non-tertiary-educated), and the urban labour force participation for the tertiary-educated at the district level. We present the results in Table 10. Our results remain statistically significant after accounting for all of these factors related to urbanization. Hence, while villages closer to Catholic missionaries were closer to urban centres, urbanization does not seem to drive the correlation between tertiary education and village prosperity.

[23] 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.

Table 9: Robustness, IV results: alternative channels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel (a): First stage: share tertiary and above (pop 25+)									
Mean sub-district 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 and 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-five mortality rate	Share Irrigated Land	Violent crimes per capita	Confidence panchayats	School index	Primary health centre	Maternity Centre	Post office	Drainage system
Anderson–Rubin (AR) test statistic $\chi^2(1)$	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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

Table 10: Robustness, IV estimates: control for factors related to urbanization

	(1)	(2)	(3)	(4)
Panel (a): First stage: share tertiary and above (pop 25+)				
Mean sub-district 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 and above (pop 25+)	0.0649**	0.0615**	0.0608***	0.0681**
	(0.0308)	(0.0307)	(0.0210)	(0.0270)
Urbanization control	District urbanization rate	Sub-district urbanization rate	Relative urban wage tertiary vs non-tertiary education	Labour force participation tertiary educated (urban)
Anderson–Rubin (AR) test statistic $\chi^2(1)$	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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

5 Mechanisms

How does having tertiary-educated members affect the village income? To answer this question one would need to construct village-level accounts, and to decompose total village income into its constituent sources. However, such accounts do not exist for rural areas. Hence we look at occupation profiles and, where possible, income in particular activities of interest to give some evidence on the possible mechanisms. In particular, we describe the two potential channels to explain the mechanisms behind the same.

5.1 Agriculture

A large part of the rural economy is in agriculture, so we start by exploring the impact of tertiary education on the agriculture sector. Extant literature on the impact of human capital on agricultural productivity is ambiguous—while early literature argued that education was important for agricultural development (Schultz 1964), recent cross-country evidence has been mixed. While some studies find no, or sometimes even negative, impact (Craig et al. 1997; Vollrath 2007), others find strong positive impacts (Foster and Rosenzweig 1996; Reimers and Klasen 2013). Such positive impacts are mediated through enhancement of decision-making skills, access to information regarding inputs and prices, adopting promising new technologies faster, and a general preferences for riskier production technology typically promising higher returns (Asadullah and Rahman 2009).

To explore whether tertiary education has an impact on agricultural income and the various mechanisms pointed out in the literature, we move to using micro-data.[24] We use the data collected in 2013 by the NSSO as part of the India-Situation Assessment Survey of Agricultural Households. We focus on 29,652 households that are involved in cultivation. Following Deolalikar (1981), we measure agricultural income by the gross value of agricultural produce per unit of cultivated land.[25] Further, while the survey does not have detailed information on the technology adopted, it asks households if they have access to technical advice regarding agriculture. We use access to such advice as a proxy for access to information about technology. Further, we also measure share of crops cultivated by the household that are neither cereal crops nor pulses as measure of crop diversification (almost all farmers grow cereals). The literature has found robust evidence on the potential of crop diversification to lower the downside risks of farming (Bezabih and Sarr 2012; Bozzola et al. 2018; Bozzola and Smale 2020; Di Falco and Chavas 2009; Di Falco et al. 2011).

The gross value of crops, amount of land cultivated, and access to technical advice are measured at the household level, but the relevant level of education is measured at the individual level. This implies that the information on education has to be aggregated to the household level. We consider as our independent variable a dummy variable that indicates the presence of a tertiary-educated member in the household. Thus we are able to look at the impact of the tertiary-educated working in agriculture (24 per cent of tertiary-educated individuals are ‘market oriented skilled agriculture and fishery’ workers) as well as those who may not be working in agriculture but may affect it by bringing to their household information on markets, technology, and prices.

[24] There is no large representative village-level data on agriculture incomes or agricultural practices, which is 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 regard to impact of education on agricultural productivity are because of the choice of wrong education variables like enrolment and literacy; if one uses education attainment instead, then both the micro and macro returns to education are positive. This is consistent with the variable we use to measure human capital.

[25] 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 labour time goes into cultivation, we cannot calculate the agricultural output per worker.

Analogous to the regressions above, we identify the causal impact of tertiary education using an IV strategy. However, in the survey data for agriculture, the lowest level of geographical unit that one can identify is the district. Hence, to use an instrument analogous to the one used in the village regressions, we use the log of the average distance to a Catholic missionary, but now this is averaged across villages for the district. Given that now our instrument varies at the district level, we cluster our standard errors at the district level. In addition to the district-level control variables used in the village regressions, we include the following household-specific controls: mean age of the household, dummy variables for belonging to a 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 for land usage).

Results from this empirical exercise, reported in Table 11, show that having a tertiary-educated individual positively impacts the gross income per acre of land for agricultural households (column 1). The first-stage F statistic is low and therefore we also report the results of the Anderson–Rubin test for weak instruments. Further we report the weak instrument robust Anderson–Rubin confidence intervals. The Anderson–Rubin test statistic rejects a weak instrument problem. Further, the 95 per cent Anderson–Rubin weak instrument robust confidence intervals in each IV regression suggest that the coefficient of interest in each case is bounded away from zero. In addition, we find that households with tertiary-educated members tend to have better access to technical advice (column 2). There is also some suggestive evidence that households with such tertiary-educated members diversify their crop choice, with the Anderson–Rubin test confidence interval bounded away from zero (column 3).

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 test statistic $\chi^2(1)$	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

Note: 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).

Source: authors' calculations.

Since survey based data on gross income from agriculture can be subject to measurement error and because our results are at the micro level, we also check whether satellite-based measures confirm the general finding at a more aggregate level. For this we use data on net primary product, which is a satellite-based measure of agriculture productivity (Zaveri et al. 2020). These data are gridded, where grids cover many villages—hence we calculate this measure for each sub-district of India and use its log as the dependent variable for our regressions. Our regression model is run at the sub-district level with the average rural tertiary completion rate calculated at the sub-district level. The current controls are modified to be at the sub-district level (SC and ST population, total population) along with the other district-level controls that we use in the village regressions. Given that the estimation is run at the sub-district level, we use, as in the village regressions, the log of average distance to a Catholic missionary at the sub-district level as the relevant instrument. The results in Table 12 confirm the micro-results; sub-districts with higher rural tertiary completion rates have a positive causal association with agricultural productivity, as measured by net primary product.

Table 12: Mechanism, IV estimates: net primary productivity (log)

	(1)	(2)	(3)	(4)	(5)
Panel (a): First stage: share tertiary and above (pop 25+)					
Mean sub-district 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 and 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 test statistic $\chi^2(1)$	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

Note: 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).

Source: authors' calculations.

The impact of tertiary-educated members on village prosperity is also confirmed by heterogeneity analysis where we look at the interaction of tertiary education and agriculture-friendly climate and soil conditions. To do so, in columns 1 and 2 of Table 13 we consider the districts where the soil is of good and bad quality, respectively.[26] We see that the results are significant only for districts where the quality of soil is better. We conduct an analogous analysis with rainfall and find that the results are significant when the district-level rainfall is good. These findings show that agriculture has a key role to play in how tertiary education affects village prosperity.

[26] We define a district to have good-quality soil if the proportion of sandy soil is less than the median proportion of sandy soil for the country (around 2 per cent). When the proportion of sandy soil is more the median proportion of sandy soil, we label it a bad-quality soil district.

Table 13: Heterogeneity, IV estimates

	(1)	(2)	(3)	(4)
Panel (a): First stage: share tertiary and above (pop 25+)				
Mean sub-district 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 and 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, ∞]	[-∞, ∞]	[-∞,-0.0495] ∪ [0.1119, ∞]	[0.0589, 0.1213]
Anderson–Rubin test statistic $\chi^2(2)$	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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

These results are interesting and new. Most of the existing literature finds that the returns to agriculture from human capital are high for primary and secondary education. There are three marked differences in understanding these results when compared to the literature. First, most of the seminal papers in the context of India (Foster and Rosenzweig 1996) are in the context of the Green Revolution—tertiary education completion rates were less than 1–2 per cent in the 1970s and 1980s and were largely concentrated in urban areas. Second, early work by Schultz (1975) found that college education did have a positive impact on agriculture, but in contexts where agriculture was not traditional (for similar evidence on a developing country, see Pudasaini (1983)). Since Indian agriculture has seen increasing mechanization over the years, this result is consistent with the earlier evidence. Third, most papers look at the impact of the human capital of farmers on productivity. As pointed out above, we look at the impact of having a tertiary-educated member in the household. Recall that only one-quarter of such individuals are actually engaged in agriculture. This points to possible externalities of having tertiary-educated people in households. Such members are more likely to be aware of external markets, prices, and newer technologies. Our work therefore suggests a new channel through which human capital externalities can impact agriculture.

To summarize, micro-results find impacts of tertiary education on agriculture. While we are not able to explain how much of the village domestic product is explained by this channel, our results show that it is driven, to an extent, by production activities within the village. This also suggests that tertiary education can affect the structural transformation process. In addition to the effect driven by the services sector, the evidence found in this paper shows that tertiary education can also influence a structural transformation process through the agricultural sector.

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 labour force would reallocate labour from the farm to the non-farm sector (Caselli and Coleman II 2001) and lead to a rise in income. However, across 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 are looking at the tertiary-educated living within the village, the rise of prosperity hints at other processes at play. To delve deeper, we again use micro-data for rural India. Our analysis relies on employment data collected by the NSSO over 12 months spanning 2009–10. We focus on 125,023 adults aged 25 and above living in rural India. Occupations are classified into farmers, those running non-farm businesses (also called the self-employed), those working as agricultural labour, those working for daily wages in the non-farm sector, and those with regular wage jobs in the non-farm sector. While the income profiles of all occupations are not available in one representative survey, there is consensus that, on average, a regular wage job (also called a salaried job) is typically the highest-paying profession (Fulford 2014). There is, however, a problem of interpreting the higher pay in regular jobs as the value of the marginal product. It is well known that in developing countries an inefficient public sector is an important contributor to regular jobs (Pritchett 2001). Private regular jobs are therefore better markers of productive jobs. Given this, we examine the impact of tertiary education on the probability of regular wage jobs, and then look at the probability of public and private regular jobs separately.

We estimate the impact of being tertiary-educated at the individual level. Hence, 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 labour force. As before, due to lack of geographical identifiers below the district level, we use as the instrument the log of distance to a Catholic missionary, averaged over villages in a district.[27] Panel (a) of Table 14 suggests that while our first stage is valid, our instrument may be weak. Hence, analogous to the above, we test for weak instruments and report the 95 per cent weak instrument robust confidence interval.

The second-stage results in panel (b) point to an impact of tertiary education on obtaining a private regular job. A similar impact is not seen for the public sector jobs and therefore not for regular jobs as a whole. This is not entirely surprising as the public sector hires adults at various levels of education (matriculation, after higher secondary, and after graduation). The overall insignificant impact on regular jobs is driven by a large share of such jobs being in the public sector. However, the significance of private jobs points to an income increase that is linked to productivity.

Given the size of the village economy, there is likely to be a relatively small supply of regular wage jobs—most of such jobs in the village are public. Commuting for work, however, is very common in rural areas and is much more popular than migration (Sharma and Chandrasekhar 2014). Hence, we look at whether being tertiary-educated increases the probability of individuals commuting daily to urban areas for work. The results in columns 4 and 5 show that while being tertiary-educated does not increase the probability of commuting to urban areas for work, it raises the probability of going out for private work. This suggests an interesting link between the rural and urban economy—one that is mediated by the tertiary educated.[28] It also suggests an interesting perspective on how we think of structural transformation. Given our results, 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 to urban areas for work.

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

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

Table 14: 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 test statistic $\chi^2(1)$	6.14**	0.90	1.04	1.07	4.54**	10.78***
Weak IV robust 95% confidence interval	[0.1600, ∞]	[-∞, 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 and household controls	YES	YES	YES	YES	YES	YES
Geographical characteristics	YES	YES	YES	YES	YES	YES
Historical variables	YES	YES	YES	YES	YES	YES

Note: robust standard errors are in parentheses and clustered at the district level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Individual and 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).

Source: authors' calculations.

However, all the impact from tertiary-educated individuals does not emanate from outside the village. Our previous subsection has shown the positive impacts on agriculture of having tertiary-educated individuals. We test here if being tertiary-educated has a causal impact on being a skilled agriculture farmer/worker. The results in column 6 show that this is indeed the case.

6 Conclusion and discussion

Tertiary education is growing at a rapid pace in India. This growth is not restricted to urban places; rural areas have also witnessed an increase in their highly qualified populations in recent decades. While the literature has focused on the relation between the increase in tertiary-educated individuals, the growth of the service sector, and increased in urbanization rates, little is known about whether tertiary-educated individuals have promoted rural development and, if so, through which channels this has happened.

In this paper we use census data on Indian villages and analyse the role of tertiary-educated individuals in rural prosperity. We address the endogeneity of tertiary education with the location of Catholic missionaries in 1911. 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 nighttime light density. A 1 per cent increase in the proportion of the population with tertiary education increases log mean consumption per capita by 7 per cent. 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 other institutions and interventions—health (proxied by under-five mortality rates), crime rates, irrigation, confidence in local village councils, village public amenities such as primary health centres, schools, and drainage.

Delving into what mechanisms drive the link between tertiary education and village prosperity, we find an important role for agriculture. The presence of tertiary-educated individuals in a household raises agriculture gross income and increases crop diversification. Households with such educated people are more likely to have access to technical advice on agriculture. These results are confirmed by satellite data that measure net primary product of an area. The important role for agriculture as a mechanism is buttressed by the fact that the results are more robust for districts with good-quality soil and rainfall. Further, we look into the impact of tertiary education on occupations and find that tertiary education enables adults to take up skilled occupations—in both agriculture and the private job market.

The stylized models of structural transformation associate higher education with migration to cities to take up non-farm jobs. Thus, higher education is often associated with increasing urbanization. Village prosperity can be achieved in such contexts through remittance transfers. However, our results show that if the tertiary-educated live in rural areas, as is increasingly the case in India, they can affect the village economy by bringing their skills and knowledge of markets/technology to bear in agriculture and by engaging in productive private sector jobs in nearby commutable areas. Our results call, therefore, for a reassessment of how we think about the role of higher education in structural transformation.

References

Acemoglu, D., F.A. Gallego, and J.A. Robinson (2014). 'Institutions, Human Capital, and Development'. *Annual Review of Economics*, 6(1): 875–912.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). 'Ethnic Inequality'. *Journal of Political Economy*, 124(2): 428–88. https://doi.org/10.1086/685300

Asadullah, M.N., and S. Rahman (2009). 'Farm Productivity and Efficiency in Rural Bangladesh: The Role of Education Revisited'. *Applied Economics*, 41(1): 17–33. https://doi.org/10.1080/00036840601019125

Asher, S., and P. Novosad (2020). 'Rural Roads and Local Economic Development'. *American Economic Review*, 110(3): 797–823. https://doi.org/10.1257/aer.20180268

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (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*, 35(4): 845–71. https://doi.org/10.1093/wber/lhab003

Baum-Snow, N., and R. Pavan (2012). 'Understanding the City Size Wage Gap'. *The Review of Economic Studies*, 79(1): 88–127. https://doi.org/10.1093/restud/rdr022

Bellemare, M.F., and C.J. Wichman (2020). 'Elasticities and the Inverse Hyperbolic Sine Transformation'. *Oxford Bulletin of Economics and Statistics*, 82(1): 50–61. https://doi.org/10.1111/obes.12325

Bezabih, M., and M. Sarr (2012). 'Risk Preferences and Environmental Uncertainty: Implications for Crop Diversification Decisions in Ethiopia'. *Environmental and Resource Economics*, 53: 483–505. https://doi.org/10.1007/s10640-012-9573-3

Birthal, P.S., D. Roy, and D.S. Negi (2015). 'Assessing the Impact of Crop Diversification on Farm Poverty in India'. *World Development*, 72: 70–92. https://doi.org/10.1016/j.worlddev.2015.02.015

Bozzola, M., E. Massetti, R. Mendelsohn, and F. Capitanio (2018). 'A Ricardian Analysis of the Impact of Climate Change on Italian Agriculture'. *European Review of Agricultural Economics*, 45(1): 57–79. https://doi.org/10.1093/erae/jbx023

Bozzola, M., and M. Smale (2020). 'The Welfare Effects of Crop Biodiversity as an Adaptation to Climate Shocks in Kenya'. *World Development*, 135: 105065. https://doi.org/10.1016/j.worlddev.2020.105065

Calvi, R., L. Hoehn-Velasco, and F.G. Mantovanelli (2022). 'The Protestant Legacy Missions, Gender, and Human Capital in India'. *Journal of Human Resources*, 57(6): 1946–80. https://doi.org/10.3368/jhr.58.2.0919-10437R2

Calvi, R., and F.G. Mantovanelli (2018). 'Long-Term Effects of Access to Health Care: Medical Missions in Colonial India'. *Journal of Development Economics*, 135: 285–303. https://doi.org/10.1016/j.jdeveco.2018.07.009

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. https://doi.org/10.1086/321015

Casey, K., R. Glennerster, E. Miguel, and M. Voors (2023). 'Skill Versus Voice in Local Development'. *The Review of Economics and Statistics*, 105(2): 311–26. https://doi.org/10.1162/rest_a_01082

Castelló-Climent, A., L. Chaudhary, and A. Mukhopadhyay (2018). 'Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India'. *The Economic Journal*, 128(616): 3039–75. https://doi.org/10.1111/ecoj.12551

Castelló-Climent, A., and A. Mukhopadhyay (2013). 'Mass Education or a Minority Well Educated Elite in the Process of Growth: The Case of India'. *Journal of Development Economics*, 105: 303–20. https://doi.org/10.1016/j.jdeveco.2013.03.012

Chanda, A., and C.J. Cook (2022). 'Was India's Demonetization Redistributive? Insights from Satellites and Surveys'. *Journal of Macroeconomics*, 73: 103438. https://doi.org/10.1016/j.jmacro.2022.103438

Craig, B.J., P.G. Pardey, and J. Roseboom (1997). 'International Productivity Patterns: Accounting for Input Quality, Infrastructure, and Research'. *American Journal of Agricultural Economics*, 79(4): 1064–76. https://doi.org/10.2307/1244264

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): 275–79. https://doi.org/10.2307/1239565

Di Falco, S., and J.-P. Chavas (2009). 'On Crop Biodiversity, Risk Exposure, and Food Security in the Highlands of Ethiopia'. *American Journal of Agricultural Economics*, 91(3): 599–611. https://doi.org/10.1111/j.1467-8276.2009.01265.x

Di Falco, S., M. Veronesi, and M. Yesuf (2011). 'Does Adaptation to Climate Change Provide Food Security? A Micro-Perspective from Ethiopia'. *American Journal of Agricultural Economics*, 93(3): 829–46. https://doi.org/10.1093/ajae/aar006

Elbers, C., J.O. Lanjouw, and P. Lanjouw (2003). 'Micro-Level Estimation of Poverty and Inequality'. *Econometrica*, 71(1): 355–64. https://doi.org/10.1111/1468-0262.00399

Finlay, K., and L.M. Magnusson (2009). 'Implementing Weak-Instrument Robust Tests for a General Class of Instrumental-Variables Models'. *The Stata Journal*, 9(3): 398–421. https://doi.org/10.1177/1536867X0900900304

Foster, A.D., and M.R. Rosenzweig (1995). 'Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture'. *Journal of Political Economy*, 103(6): 1176–209. https://doi.org/10.1086/601447

Foster, A.D., and M.R. Rosenzweig (1996). 'Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution'. *American Economic Review*, 86(4): 931–53.

Fulford, S. (2014). 'Returns to Education in India'. *World Development*, 59: 434–50. https://doi.org/10.1016/j.worlddev.2014.02.005

Gennaioli, N., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2013). 'Human Capital and Regional Development'. *The Quarterly Journal of Economics*, 128(1): 105–64. https://doi.org/10.1093/qje/qjs050

Gibson, J. (2021). 'Better Night Lights Data, for Longer'. *Oxford Bulletin of Economics and Statistics*, 83(3): 770–91. https://doi.org/10.1111/obes.12417

Gibson, J., S. Olivia, G. Boe-Gibson, and C. Li (2021). 'Which Night Lights Data Should We Use in Economics, and Where?'. *Journal of Development Economics*, 149: 102602. https://doi.org/10.1016/j.jdeveco.2020.102602

Gollin, D., D. Lagakos, and M.E. Waugh (2014). 'Agricultural Productivity Differences across Countries'. *American Economic Review*, 104(5): 165–70. https://doi.org/10.1257/aer.104.5.165

Harris, J.R., and M.P. Todaro (1970). 'Migration, Unemployment and Development: A Two-Sector Analysis'. *American Economic Review*, 60(1): 126–42.

Hazra, C. (2001). 'Crop Diversification in India'. In M.K. Papademetriou and F.J. Dent (eds), *Crop Diversification in the Asia-Pacific Region*. Rome: FAO.

Henderson, J.V., A. Storeygard, and D.N. Weil (2012). 'Measuring Economic Growth from Outer Space'. *American Economic Review*, 102(2): 994–1028. https://doi.org/10.1257/aer.102.2.994

Herrendorf, B., and T. Schoellman (2018). 'Wages, Human Capital, and Barriers to Structural Transformation'. *American Economic Journal: Macroeconomics*, 10(2): 1–23. https://doi.org/10.1257/mac.20160236

Hicks, J.H., M. Kleemans, N.Y. Li, and E. Miguel (2017). 'Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata'. Working Paper 23253. Cambridge, MA: NBER. https://doi.org/10.3386/w23253

Joshi, P.K., A. Gulati, P.S. Birthal, and L. Tewari (2004). 'Agriculture Diversification in South Asia: Patterns, Determinants and Policy Implications'. *Economic and Political Weekly*, 39(24): 2457–67.

Lewis, W.A. (1954). 'Economic Development with Unlimited Supplies of Labour'. *The Manchester School*, 22(2): 139–91. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x

Maloney, W.F., and F. Valencia Caicedo (2017). 'Engineering Growth: Innovative Capacity and Development in the Americas'. Available at: https://blogs.lse.ac.uk/latamcaribbean/2018/05/15/engineering-growth-innovative-capacity-and-development-in-the-americas (accessed 26 June 2023).

Mantovanelli, F. (2014). 'The Protestant Legacy: Missions and Literacy in India'. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2413170 (accessed 26 June 2023).

McMillan, M., D. Rodrik, and C. Sepulveda (2017). 'Structural Change, Fundamentals and Growth: A Framework and Case Studies'. Working Paper 23378. Cambridge, MA: NBER. https://doi.org/10.3386/w23378

Meisenzahl, R.R., and J. Mokyr (2011). 'The Rate and Direction of Invention in the British Industrial Revolution: Incentives and Institutions'. In J. Lerner and S. Stern (eds), *The Rate and Direction of Inventive Activity Revisited*. Chicago, IL: University of Chicago Press.

Michalopoulos, S., and E. Papaioannou (2013). 'Pre-Colonial Ethnic Institutions and Contemporary African Development'. *Econometrica*, 81(1): 113–52. https://doi.org/10.3982/ECTA9613

Moretti, E. (2004). 'Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions'. *American Economic Review*, 94(3): 656–90. https://doi.org/10.1257/0002828041464623

Porzio, T., F. Rossi, and G. Santangelo (2022). 'The Human Side of Structural Transformation'. *American Economic Review*, 112(8): 2774–814. https://doi.org/10.1257/aer.20201157

Porzio, T., and G. Santangelo (2017). 'Structural Change and the Supply of Agricultural Workers'. Discussion Paper. London: CEPR.

Pritchett, L. (2001). 'Where Has All the Education Gone?'. *World Bank Economic Review*, 15(3): 367–91. https://doi.org/10.1093/wber/15.3.367

Pudasaini, S.P. (1983). 'The Effects of Education in Agriculture: Evidence from Nepal'. *American Journal of Agricultural Economics*, 65(3): 509–15. https://doi.org/10.2307/1240499

Reimers, M., and S. Klasen (2013). 'Revisiting the Role of Education for Agricultural Productivity'. *American Journal of Agricultural Economics*, 95(1): 131–52. https://doi.org/10.1093/ajae/aas118

Roca, J.D.L., and D. Puga (2017). 'Learning by Working in Big Cities'. *The Review of Economic Studies*, 84(1): 106–42. https://doi.org/10.1093/restud/rdw031

Sachs, J. (2003). 'Institutions Don't Rule: Direct Effects of Geography on Economic Development'. Working Paper 9490. Cambridge, MA: NBER. https://doi.org/10.3386/w9490

Schultz, T.W. (1964). 'Changing Relevance of Agricultural Economics'. *Journal of Farm Economics*, 46(5): 1004–14.

Schultz, T.W. (1975). 'The Value of the Ability to Deal with Disequilibria'. *Journal of Economic Literature*, 13(3): 827–46.

Sharma, A., and S. Chandrasekhar (2014). 'Growth of the Urban Shadow, Spatial Distribution of Economic Activities, and Commuting by Workers in Rural and Urban India'. *World Development*, 61: 154–66. https://doi.org/10.1016/j.worlddev.2014.04.003

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL: 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): 1825–83. https://doi.org/10.1093/qje/qjv025

Taylor, J.E., and P.L. Martin (2001). 'Human Capital: Migration and Rural Population Change'. *Handbook of Agricultural Economics*, 1: 457–511. https://doi.org/10.1016/S1574-0072(01)10012-5

Vollrath, D. (2007). 'Land Distribution and International Agricultural Productivity'. *American Journal of Agricultural Economics*, 89(1): 202–16. https://doi.org/10.1111/j.1467-8276.2007.00973.x

Wu, K., and X. Wang (2019). 'Aligning Pixel Values of DMSP and VIIRS Nighttime Light Images to Evaluate Urban Dynamics'. *Remote Sensing*, 11(12): 1463. https://doi.org/10.3390/rs11121463

Young, A. (2013). 'Inequality, the Urban–Rural Gap, and Migration'. *The Quarterly Journal of Economics*, 128(4): 1727–85. https://doi.org/10.1093/qje/qjt025

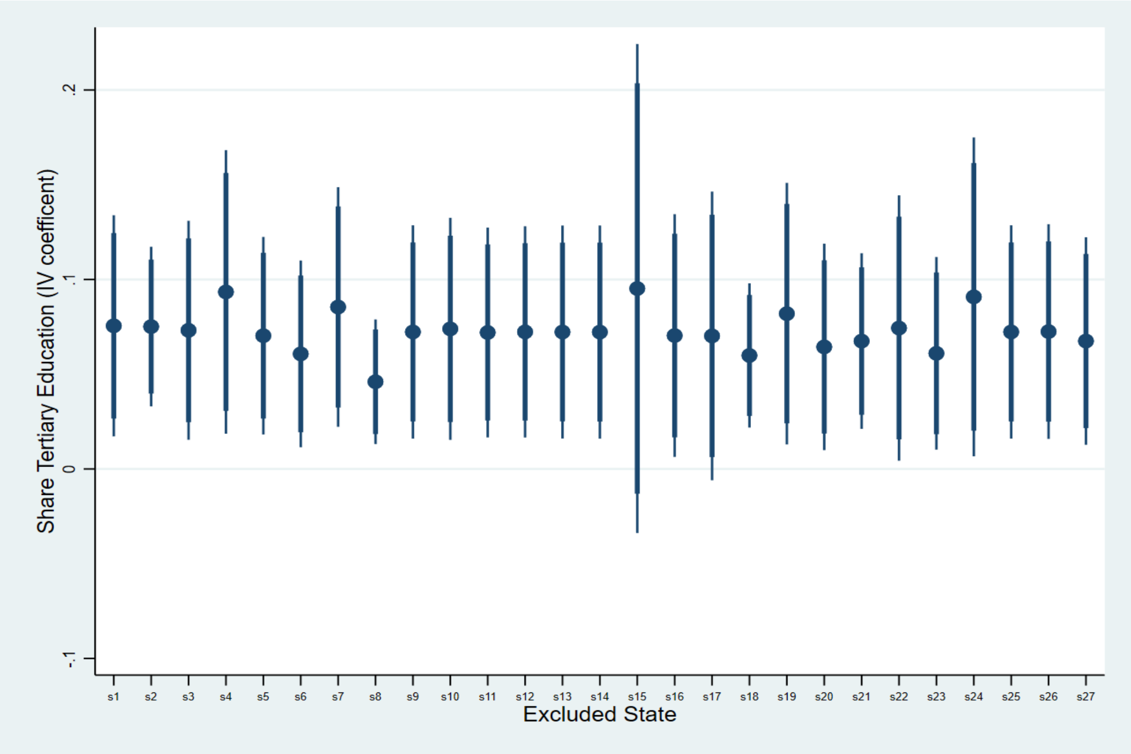
Zaveri, E., J. Russ, and R. Damania (2020). 'Rainfall Anomalies Are a Significant Driver of Cropland Expansion'. *Proceedings of the National Academy of Sciences*, 117(19): 10225–33. https://doi.org/10.1073/pnas.1910719117

Appendix A: Village and town description

According to the 2011 Census of India, around 68 per cent 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+. As per the 2011 Census, we do not 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 per cent of the male working population engaged in non-agricultural pursuits, and a population density of at least 400/km^2 (1,000/mi^2). Towns in India usually have basic infrastructure such as 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.

Appendix B: Tables and figures

Figure B1: Robustness, IV estimates: excluding one state



Note: 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.

Source: authors' calculations.

Table B1: Robustness, including lower level of education

	(1)	(2)	(3)	(4)
Panel (a): First stage: share tertiary and above (pop 25+)				
Mean sub-district 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 and 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 test statistic $\chi^2(1)$			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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.

Table B2: OLS results, Catholic missionary on alternative channels

	Under-five mortality rate	Share irrigated land	Violent crimes per capita	Confidence panchayats	School index	Primary health centre	Maternity centre	Post office	Drainage system
Mean sub-district 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

Note: robust standard errors are in parentheses and clustered at the sub-district 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 km^2 (log). Historical variables: fraction of Brahman 1931, fraction of tribal 1931, fraction of urban 1931, Princely State (dummy), and rail (1909).

Source: authors' calculations.