The Protestant Legacy: Missions and Literacy in India

Federico G. Mantovanelli[†]

Job Market Paper

This Version: December, 2013

Abstract

As an important expression of culture, religion represents a possible fundamental source of economic and social outcomes. This paper investigates how the historical development of Protestantism may contribute to explain current literacy disparities in India. In order to enable everyone to read the Bible by themselves, Protestants have always stressed the importance of promoting universal literacy. Combining information about the spatial distribution of Protestant missions in India at the end of the nineteenth century with contemporary district-level data, this paper documents a strong long-term relationship between the historical exposure to Protestant missions and current literacy. This pattern does not depend on either local geographic characteristics or the level of historical development of the districts. By exploiting only the variation within groups of geographically contiguous districts and using historical Catholic missions as control group, I verify that this relationship is not driven by unobserved characteristics that may affect both current literacy outcomes as well as the missionaries' location decisions.

[†]✉ Department of Economics, Boston College, Chestnut Hill, MA 02467. E-mail: federico.mantovanelli@bc.edu. Web: https://sites.google.com/site/fedmantovanelli/.

I thank Scott Fulford, Fabio Schiantarelli and Mathis Wagner for their constant support and help. I am grateful to Samson Alva, Andrew Beauchamp, Matteo Cacciatore, Rossella Calvi, Donald Cox, Katherine Griwert and Nidhiya Menon for their valuable comments and suggestions. I also want to thank seminar participants at the BC-BU Applied Micro Meeting, the Boston College Applied Micro Workshop and the Boston College Dissertation Workshop. I am indebted to the staff of the Lamont library (Harvard University) for granting me access to the volumes of the 1901 Census of India. All errors are mine.

1 Introduction

Human capital is a key determinant of cross-country and regional development (Hanushek and Woessmann [2012]; Gennaioli *et al.* [2012]). The economic growth of many emerging countries and their ability to reduce the gap with the rich economies is due, among other factors, to their rapidly growing literacy rates. It is not surprising, therefore, that the promotion of literacy has become a central issue in the policy agenda of both national governments and international organizations. India represents a vivid example of this trend. During the last decades, several nationwide programs have been implemented to improve the status of literacy and education.[1] Although these interventions have increased literacy in the country, wide and significant disparities still exist across both geographic areas and social groups. Since differences in literacy and education lead economic and social inequalities, it is important to understand how they first emerge and persist over time.[2] Inspired by a growing empirical literature that looks at how history and historical institutions contribute to the determination of contemporary outcomes, the present paper exploits the spatial distribution of Protestant missions in colonial India in order to explain current literacy levels.[3]

As an important expression of culture, religion is viewed as a possible fundamental source of economic and social outcomes (Barro and McCleary [2003, 2005]). Protestantism, in particular, has a direct link to the advancement of human capital. The principle of "Sola Scriptura", one of the pillars of the Protestant theology, implies that one can understand God's word only by reading the Bible.[4] Quite obviously, to be able to read and understand the Scriptures, one first needs to be able to read. Consequently, Protestants have always stressed the importance of promoting universal literacy and education in order to give everyone the capability of reading and understanding the Bible.

Consistent with that view, in this paper I explore the long-term relationship between the exposure to historical Protestant missions and current differences in literacy outcomes across districts in India. To this aim I construct a novel data set that identifies the exact location of Protestant missions operating in India between the end of the nineteenth and the beginning of the twentieth century. I then link this information to contemporary and historical district-level data.

[1] Between 2009 and 2012, $1.5 billion have been allocated by the Indian government to the "Saakshar Bharat Mission" (India Literate Mission), a project aimed at promoting adult literacy and reaching out those who did not access a complete education (the program was first launched in 1988 under the name of National Literacy Mission). A similar initiative, the "District Primary Education Program", has been financed, since the late 1990s, by the World Bank to facilitate India's efforts to achieve universal primary education.

[2] Card (1999) and Acemoglu and Angrist (2000) show how differences in education lead to disparities in labor market outcomes. Cutler and Lleras-Muney (2006) find that education affects health outcomes. Kam and Palmer (2008) and Berinsky and Lenz (2011) present evidence of a direct effect of education on the individual level of political participation.

[3] The seminal works are Engerman and Sokoloff (2002), looking at the importance of factor endowments and colonial rule for the subsequent development of colonies in America; and Acemoglu et al. (2001), examining the effect of colonial rule on the development of institutions in Africa. Since those early contributions, several works have identified historical events as an important source for current economic and social outcomes such as trust (Nunn and Wantchekon [2011]), the supply of public goods (Banerjee and Iyer [2005]), consumption levels (Dell [2010]). For a more extensive and detailed review of this literature, please refer to Nunn (2009)

[4] Protestants believe that the Bible is self-authenticating and clear to the rational reader who does not need the intermediation of a cleric member: "*The Christian reader should make it his first task to seek out the literal sense, as they call it. For it alone is the whole substance of faith and Christian theology*" (Martin Luther, as reported in Wood [1969]). Unlike the Catholic and the Orthodox Churches, the Protestant Reformation considers the Scripture as the ultimate authority of the Christian doctrine: "*A simple layman armed with the Scripture is greater than the mightiest pope without it*" (Martin Luther, as reported in Bainton [1950]).

Relying on this novel data set, I test whether districts that have been exposed to Protestant missions more than one century ago are today characterized by better literacy outcomes. The answer is affirmative. Protestantism has a strong and positive long-term association with the total population's current literacy rate. Historians highlight the important role of Protestant missionaries in promoting literacy among traditionally disadvantaged groups in India: women and low castes. Consistently, I find that women and low castes members living in "Protestant" districts display significantly higher literacy rates relative to the same social groups living in "non-Protestant" areas.

These findings do not necessarily imply a causal relationship. One may be concerned that they are only the consequence of a spurious correlation due to either reverse causality or omitted variables. The former is likely not to be an issue in this context since historical missions, being established more than a hundred years ago, are clearly predetermined with respect to the outcome variables. Nonetheless, there may still be concerns related to omitted variables. Given that Protestant missionaries did not locate randomly in India, it is possible that districts' characteristics affect current literacy outcomes as well as the location decisions of the early missionaries. This would inherently prevent a causality interpretation of the previous results. To address this issue, I pursue a number of different strategies.

The first strategy is to directly control for those factors that most likely affected the missionaries' location decisions. A set of geographic variables is included to properly account for geographic heterogeneity. Using district level information from the 1901 census of India, I also add an extensive set of historical characteristics in order to measure the districts' level of social and economic development at a time contemporaneous to the diffusion of Protestant missions in India. Following Altonji et al. (2005), I calculate how much greater the influence of unobservables needs to be, relative to observable factors, in order to explain away the positive relationship between Protestant missions and literacy. I show that the influence of unobservables would have to be more than twice as great as observable factors. Therefore, it is unlikely for unobserved heterogeneity to be driving my findings.

The second strategy follows Banerjee and Iyer (2005) and Michalopoulos and Papaioannou (2013) to show that the relationship between historical missions and current literacy continues to hold within groups of neighboring districts that happen to differ only in terms of their exposure to Protestant missions.

My third strategy uses historical Catholic missions as control group. Over the course of the nineteenth century, the British Crown adopted a neutral policy regarding missionary activities in its colonial territories. Groups of different Christian denominations were free to operate with no specific restrictions. It is therefore plausible to assume that factors that drove the location decisions of Protestant missionaries are common to Catholic missionaries, too. When I account for the presence of historical Catholic missions, I find no effect of the Catholic exposure on current literacy. That suggests that it is not unobserved characteristics of the districts themselves, but instead the activities of the early

Protestant missionaries that drive the results.

My most conservative estimates indicate that the exposure to Protestant missions determines an increase of 2.5 percentage points of the total literacy rate and of 3 percentage points of the female literacy rate. The literacy gender parity index (measuring the degree of literacy equality between men and women) and the literacy rate of low castes individuals are, respectively, 2 points and 2.5 percentage points higher in "Protestant" districts relative to "non-Protestant" areas.

After establishing that Protestant missions have a positive long-term effect on current literacy outcomes, I analyze the potential channels through which this effect may operate. Missions may have had a positive contribution by directly increasing the local supply of education through the foundation of schools. At the same time, missions may have laid the ground for the future development of the districts: by maintaining contact with their countries of origin and constant access to supplies from Europe, for example, missionaries may have improved a district's openness to trade, fostering its economic development. I show that even when controlling for the district's diffusion of schools and its current degree of development, a significant portion of the missions-literacy relationship remains unaccounted for. One possible explanation is that part of this long-term effect is demand driven, that is due to a persistent change in people's attitudes toward education.[5] By preaching the importance of education and the acquisition of literacy skills, the first missionaries may have changed people's views about the relevance of being educated. It is likely that this "new" attitude has been transmitted across generations so that the descendants of those in contact with the early missionaries value human capital more and are therefore more literate today.[6] I also show that this effect is not limited to the Christian community. Even when the district's proportion of Christians is accounted for, the long-term effect of Protestant missions survives suggesting that it is the result of a broader cultural change rather than just a direct religious effect.

To the best of my knowledge, this paper represents the first attempt to quantitatively assess the long-term effect of Protestant missionaries on human capital in India. The idea that religious principles and ideas can explain economic outcomes dates back to Max Weber.[7] Recent works have focused on the relationship between religion and educational outcomes in several different settings. Becker and Woessmann (2009) use the population share of Protestants to explain differences in human capital levels across Prussian counties during the nineteenth century. The present paper examines the long-term (rather than contemporaneous) effect of Protestantism. Gallego and Woodberry (2010) document a positive correlation between Protestant missionaries' activity at the beginning of the twentieth century

[5]This is in line with a process of intergenerational cultural transmission as defined by Bisin and Verdier (2000 and 2001).

[6]There is an extensive empirical literature documenting the significant and positive impact of parental (Oreopoulos [2003], Chevalier [2004], Black *et al.* [2005]) and neighborhood (Kremer [1997], Ioannides [2003]) background on individual educational achievement.

[7]In "The Protestant Ethic and the Spirit of Capitalism," Weber argues that Protestantism was critical for the development of capitalism in Western Europe. Since the diffusion of his study, various theories have been suggested to explain the relationship between Protestantism and economic development. Empirically, however, the findings have been mainly inconclusive, not being able to establish a clear direct impact of Protestantism on economic growth (see, for example, Cantoni [2010]). Recently, Basten and Betz (2013) exploit a historical quasi-natural experiment in South-Western Switzerland to examine the effect of Protestantism on current individual preferences for leisure and political attitudes.

and current educational outcomes across countries in Africa. Their cross-country analysis, however, does not allow to disentangle the effect of missions from other country-specific fundamentals of human capital. Moreover, they do not address concerns related to bias due to the possibly endogenous location of Christian missions.

More broadly, this study relates to the literature looking at the institutional origins of current economic and social outcomes. The within-country approach of the paper allows to overcome problems related to cross-country unobserved heterogeneity in line with a recent body of research (Banerjee and Iyer [2005]; Dell [2010]). Moreover, the results of the paper complement the micro findings of Nunn (2011) and Wantchekon et al. (2013) on the beneficial effect of colonial and missionary education in Africa.

The remainder of the paper is organized as follows. Section 2 gives a general overview of the current status of literacy in India and offers a synthetic historical description of the development of Christianity and the involvement of Protestant missionaries in the field of education. Section 3 presents the data sources and discusses the construction of the original data set. Section 4 presents the empirical approach. Section 5 reports the main findings and provides evidence that they are not driven by omitted variables. Section 6 shows the robustness of the results. Section 7 discusses potential channels through which the positive effect of historical missions may operate. Section 8 concludes.

2 Overview: Missionaries and Literacy in India

2.1 Protestant Missionaries in Colonial India

According to the tradition, Christianity arrives in India when St. Thomas, one of the twelve disciples of Jesus Christ, reaches the Malabar coast in A.D. 52. During his travel to southern India around the end of the thirteenth century, Marco Polo mentions the burial site of St. Thomas as an important place of pilgrimage. The history of modern Christianity in India, however, starts only with the advent of the earlier European colonizers. The Portuguese first and the Dutch later introduce their brands of Christianity in the regions they rule in Asia during the sixteenth and seventeenth century, but it is with the subsequent British domination in the eighteenth century that a new wave of missionary activity begins.

The development of missions is initially constrained by limits imposed by the *East India Company*, afraid to upset the local religious sentiments and to antagonize, in particular, the Hindu population by allowing a large Christian presence in its territories.[8] The situation changes in 1813 when, in response

[8] "*We are anxious that it should be understood that we are very far from being averse to the introduction of Christianity into India;* [...] *but we have a fixed and settled opinion that nothing could be more unwise and impolitic.*" (from a 1808 despatch from the Court of Directors of the East India Company as reported in Kaye [1859], page 513)

to public outrage in England, the *Company* is compelled to grant missionaries in its territories new freedom to proselytize (Moffett [2007]). The development of Christianity assumes a significant turn from the middle of the nineteenth century when, after the crushing of the Indian Mutiny in 1858, sovereignty of India passes from the *Company* to the British Crown. Queen Victoria promotes an official policy of religious impartiality that neither favors nor opposes the work of Christian missionaries. This determines the beginning of a period of Christian expansion with missionaries coming from different countries and becoming some of the most influential actors in the Indian sub-continent (Beach [1908]).

In no field are missionaries active more consistently than in that of education, seen as one of the main ways to introduce western morals in the Indian society.[9] The emergence of the Utilitarianism in European circles marks the beginning of a new approach of proselytization through education. The belief is gaining ground that the conversion of the Indians to Christianity would be the ultimate result of the spread of education among them,

> *He* [the student] *enters the school premises, becomes acquainted with mathematical science, with astronomy and geometry. Naturally he loses confidence in his own religion when he finds it contains so many ridiculous and impossible explanations* [...] *proving* [his] *religion itself to be false.*
>
> (Rev. Henry C. Tucker, *A Few Words on Teaching the Bible in Government Schools in India*, London, 1859, p.16)

The general consensus among missionaries is that Christianity, western scholarship and European civilization are all intertwined. This faith in the proselytizing power of education leads to the diffusion of mission-run schools that begin to spread very rapidly starting from the middle of the nineteenth century.[10] Protestant missions dominate the Indian educational scene, not merely in regard to the magnitude of their effort (at the turn of the century, nearly 70 percent of primary and secondary schools in the United Provinces of Agra and Oudh were mission-run) but also in regard to the ideas and the values inculcated through education.

The missionary involvement is even larger in the education of women. Female education has always been a central issue for the Protestant Churches. Martin Luther himself advocated the importance of offering educational opportunities not only to boys but also to girls,

> *Above all, the foremost and most general subject of study, both in the higher and the lower schools, should be the Holy Scriptures, and for young boys the Gospel. And would to God that every town had a girl's school also, in which the girls were taught the Gospel for an hour each day [...]*
>
> (Martin Luther [1520], "To the Christian Nobility of the German Nation Concerning the Reform of the Christian Estate")

[9] For a more detailed analysis of missionary education in colonial India, please refer to Mathew (1988) and Bellenoit (2007a).

[10] To say the truth, proselytization through education did not give the hoped results. Conversions, on the contrary, did more to hurt schools than to help them. A single baptism emptied a school, effectively robbing missionaries of any opportunity to come into contact with students. As an example, in the spring of 1890 at Ramsay College, Almora, one Brahmin student converted, generating an open rebellion of the Hindu community against the headmaster of the school (as reported in "North India Report", Council for World Mission, 1890).

Extending education to the female population is not only dictated by moral principles but is considered essential in order to facilitate the "westernization" of the Indian society. Since women are viewed as the source of a culture's morality, educating them represents the first step in the process of transforming the Indian national character:

> *The education of your females is the next great step to be taken in the regeneration of the Hindu character [...] Whatever there is of evil common to all human nature, it is best combated [...] by laying the foundations of a virtuous life in the early inculcation of a sound morality, and by teaching women to respect themselves by showing that by us also they are held in honour. Were it only for selfish considerations, you ought to educate your women.*
> (from the speech of Rev. J. Bethune to the students of Krishnagar College (West Bengal) in 1851 as reported in Sawage [1997])

The clear interest in the education of women shown by the Protestant missionaries strikingly contrasts with the position of the British Indian Government, whose involvement in the matter is, to say the least, quite weak during the nineteenth century. The Government's unwillingness to support and promote female education is partly explained as deference to Indian socio-religious sentiments which are unconducive to the education of girls. The government practically leaves the matter to local initiative. In a society that is truly indifferent to the education of its women, it is the Protestant missions that pioneer and operate, almost monopolistically, in the field of girls' schooling. The strategies available to missionary groups are essentially two.

On one hand missions provide female education in an institutional setting, that is thorough the establishment of females' schools: as early of 1840, for example, thirty schools for Hindu girls are opened by the Church Missionary Society in Calcutta and by mid-century the missionaries in Madras (currently known as Chennai) are instructing nearly 8,000 girls in day schools and boarding schools. Despite few exceptions, however, Protestants' efforts to promote education among women through the foundation of schools face the resistance of the indigenous families (particularly from the high castes) to send their daughters to school.[11] Hence this "institutional" approach reaches mainly the poorer and lower castes.

Apart from founding schools, a second strategy is developed as a way to reach the upper classes where female seclusion was more strictly enforced: "Zenana" education. This approach consists of sending missionary women (or the wives of men missionaries) directly into the Zenanas (the women's quarter of the house) to teach them some basic literacy and numeracy skills. Zenana education reveals to be quite successful as respectable Hindu families, contrary to the idea of exposing females to public gaze, are willing to welcome European female teachers in their own house. During the nineteenth century, the movement for home education in the Zenana develops quite rapidly and, in the second

[11] Alexander Duff, a Protestant missionary who played a large part in the development of education in India, was well aware of the difficulties of promoting education for females in the first phase of westernization: "*A generation of educated males must be the precursor of a generation of educated women*" (reported in Noel [1837]).

half of the century, the Zenana movement gradually becomes a regular and institutionalized form of education.

2.2 The Status of Literacy in Modern India

The latest Indian census (conducted in 2011) reports that about 75 percent of the total population is literate.[12] Even if this represents a six-fold improvement since India became independent in 1947, literacy in the country is still well below the world's average (about 85 percent according to U.N. estimates). Moreover, India is home to the world's largest illiterate population: about 35 percent of the world's illiterates live in the Indian sub-continent (UNESCO [2010]).[13] There are several factors that can help to explain these statistics.

Poverty is a widespread phenomenon in the country. Recent estimates show that 33 percent of India's population live below the poverty line and that 68 percent of Indians live with less than U.S.$ 2 a day.[14] Poverty acts as a deterrent to households' investment in human capital as poor families are likely not to consider education as a priority relative to other basic needs.

Social barriers may play an important role, too. Indian society is strictly segmented into a hierarchical system of castes that regulates the division of labor and power. Even if, since 1950, the Indian government has enacted and implemented many laws and social initiatives to improve the socio-economic conditions of the most discriminated groups (labeled as "Scheduled Castes" [SC] and "Scheduled Tribes" [ST]), disparities continue to be pronounced. Individuals belonging to SC remain disadvantaged with respect to many social indicators, including education: according to Jenkins and Barr (2006) more than 20 percent of the SC's children (of age between 7 and 17) never attended school and among the SC population the literacy rate (around 60 percent according to the 2011 Census) is well below the national average.

India is also afflicted by a lack of adequate school infrastructure. According to a recent survey, about 30 percent of schools do not have permanent structures and most of them are deficient in proper sanitation services (14 percent do not have drinking water and 60 percent have no separate toilets for girls). In addition to a shortage of classroom (on average 37 students per classroom in primary schools, and the student-classroom ratio rises to over 80 in some states), many schools lack an adequate number of teachers (the average teacher-pupil ratio in India is 1:42, but in several areas it goes well beyond that value).[15]

[12] Literacy, as defined in Census operations, is the ability to read and write with understanding in any language. A person who can merely read but cannot write is not classified as literate. Any formal education or minimum educational standard is not necessary to be considered literate. There is some debate in India concerning the appropriateness of the census' definition of literacy. The main criticism is that such measure does not allow to isolate English literacy (i.e. the type of literacy that mostly affects an individual's employability and therefore economic condition). Critics argue that, in this way, the census is likely to overestimate the "real" literacy status of the country.

[13] According to UNESCO (2010)'s estimates, in 2006 there were 759 million adults (15 and over yeas of age) lacking basic reading, writing and numeracy skills. Of these, 270 million resided in India.

[14] http://povertydata.worldbank.org/poverty/country/IND

[15] Source: "Elementary Education in India", Government of India, Ministry of Human Resource Development, 2011.

Literacy rates are not homogenous across India; the country presents wide disparities.

At geographic level there exist striking differences among states. On one side there are states, such as Kerala and Mizoram, that have literacy rates higher than 90 percent, not far from those seen in most developed countries. On the other side, there are states characterized by literacy outcomes that make them comparable to some of the poorest areas of sub-Saharan Africa. In the state of Bihar, for example, the literacy rate amounts to a depressing 60 percent. Moreover, even within the same state, we encounter significant heterogeneities across districts: in the state of Madhya Pradesh, for example, the literacy rate varies from 61 percent (district of Dhar) to 83 percent (district of Jabalpur).[16]

Literacy varies also between genders. Even if the literacy gap between males and females seems to have decreased in the last ten years, in 2011 we still see a twenty point difference between the literacy rate of men and that of women. In addition, about 70 percent of Indian illiterates are women. These facts can represent a potential restraint for India's success as female education has long been acknowledged to have a strong positive correlation with other dimensions of human and social development. Empirical evidence has shown how an increase in women's human capital is associated with a lower degree of inequality in labor market outcomes, lower fertility rates and improved children's health, and, more in general, with economic growth.[17]

Despite strong economic and social evidence of high returns to female education, most communities continue to underinvest in the education of girls. This can be the result of the role played by both parental attitudes and social values. Especially in poor families, whenever parents have to choose whether to send to school their son or their daughter, they tend to favor the former. Educating the son is seen as an investment since he will be responsible for caring for aging parents. On the contrary, the education of daughters is often seen as a waste of money because daughters will eventually live with their husband's families and the parents will not benefit directly from their education. Moreover sending a daughter to school also implies higher opportunity costs since girls are often employed in basic agricultural activities and domestic chores (as taking care of the younger siblings and of the elderly). Implicit social barriers may decrease the educational opportunities for girls, too. In more sexually segregated communities, often parents take their daughters out of school as soon as they reach puberty as a way to protect their "honor". As mentioned earlier, there is a shortage of teachers in India; this is particularly true for female teachers. A direct consequence is that parents may decide not to send to school their daughter if they know that she will be taught by a male teacher.

[16] Source: Census of India (2011) [http://censusindia.gov.in/2011-prov-results/data_files/mp/07Literacy.pdf]

[17] See, for example, Card (1999), Psacharopoulos and Patrinos (2004), Currie and Moretti (2003), Dollar and Gatti (1999).

3 Data

To investigate the relationship between historical Protestant missions and contemporary literacy outcomes, this paper uses an original data set consisting of both historical and current information. All the data are at district level.[18] Two main reasons determine this choice. First, there exist large inter-district disparities in educational outcomes; an analysis at a more aggregate level would ignore this important source of variation. Second, using district level data gives a larger sample size. The disadvantage of this choice is that district level data are, to say the least, not readily available (particularly historical data). The final data set consists of 515 districts for which I have complete information. The following sections provide a detailed description of the different sources used to build the novel data set at the core of the empirical analysis.

[Insert Table 1 about here]

3.1 Historical Data

Information concerning the location of Protestant missions in India between the end of the 19th century and the first decade of the 20th century comes from the "Statistical Atlas of Christian Missions" (the Atlas, hereinafter), produced in occasion of the World Missionary Conference of Edinburgh (Scotland) in 1910. To the best of my knowledge it is the first time that this source of information is used in an econometric analysis.[19] The Atlas represents a directory of missionary societies throughout the world and a statistical representation of their worldwide activities as of 1908. The data were collected from the official staff of the missionary societies, from their published reports or, in some cases, directly from the mission fields. It has to be noticed that the Atlas includes only those missions that conduct their operations among non-Christians. Only Protestant missions that operate among indigenous populations are recorded; missions meant to serve only immigrants (i.e. Western colonizers) are, therefore, excluded. The Atlas reports detailed statistics about the global operations and organization of several different Protestant Societies. Importantly for the scope of this paper, it also offers a series of very precise maps showing the geographic distribution of the missions (for an example of the maps contained in the Atlas refer to Figure 1 in Appendix B).

[Insert Figure 1 about here]

Using cartography tools in a geographic information system (ArcGIS), I superimpose those historical charts to a map of India that reports the districts' administrative boundaries as of 2001. The result of this geo-referencing procedure can be seen in Figure 2 with the little crosses to indicate the location of

[18] In 2001, India had 593 districts in divided in 28 States/Union Territories. The average district had an area of 5,000 squared kilometers and a population of 1.7 million, making it comparable to King County in the state of Washington, U.S.A. (area: 5,506 km^2; population: 1.9 million).

[19] I use the Atlas in another paper of mine (Mantovanelli [2012]) where I look at the long term effect of Christian missions on individual sexual behavior and HIV infection rates in sub-Saharan Africa.

Protestant missions as of 1908. Based on this map I create a binary variable equal to one if the district was inhabited by Protestant missions as of 1908, and zero otherwise. This variable will represent my measure of historical Protestant influence.[20]

[Insert Figure 2 about here]

The map documents how Protestant missions were active in different regions of the Indian sub-continent. South-eastern districts show a larger concentration of missions but evidence of the presence of Protestant settlements is found also in the North (especially in the area around Delhi and Agra) and in the East (particularly in the area of Calcutta and the state of Orissa).

I combine information about the Protestant settlements with historical data concerning the development of the districts. These data are primarily from the 1901 Indian census, integrated with additional information contained in other statistical sources from that same historical period.

India has a long history of census operations. For obvious reasons, such as defense, collection of revenues and taxes and employment of population in profitable trades and services, the *East India Company* was anxious to obtain reliable estimates of population in its Indian settlements. These surveys, mostly conducted during the eighteenth century, were limited, however, to the *Company*'s territories and did not offer therefore a systematic picture of the demographic and economic status of the whole Indian sub-continent. In 1801 England had begun its census series and the Parliament was aiming to use that same tool in order to ascertain the population of the colonial territories, too. In 1856, the Government of India, following the directives of the British Parliament, had entered upon a consideration of the means by which a general census of the population of India might be taken in 1861. The violent outburst of the "Indian Mutiny" (1857) led to the postponement of the census operations, that were finally undertaken between 1867 and 1872. This series, commonly known as the "Census of 1872", was not, however, a synchronous project, nor did it cover all territory possessed or controlled by the British. Though based on uniform schedules it was not centrally supervised or compiled. The first complete census of population was conducted only in 1881, on a homogenous basis throughout India, providing the most complete and continuous demographic record for any comparable population. Since then the census is being regularly conducted every ten years.

This paper uses district level information from the census of 1901 for three main reasons. First, it has a wider geographical coverage relative to the censuses of 1881 and 1891. Second, it provides more detailed information about economic activities. Third, it is closer to the period covered by the Atlas. Using data from the twenty-six volumes of the 1901 census, I construct variables that measure the historical level of development of both the territories that were under the direct British control and the so called feudatory states that were ruled by a local prince.[21]

[20] Later in the paper I will use continuous variables in order to exploit also variations in the intensity of Protestant influence. See section 6.2 for more details.

[21] Some of the "princely states" in the Bombay presidency were not surveyed in the census. For these territories I rely, whenever possible, on various District Gazetteers published between 1901 and 1910 and on the "Imperial Gazetteer of India", published in 1909.

Population density and the share of population living in urban areas are used as proxies for the level of modernization of the colonial district. In addition, the Census of 1901 divides the district's labor force in four categories: agriculture, industry, commerce and professions. As an additional proxy of the level of development of historical districts, I also include the share of labor force not employed in the agricultural sector. The set of historical characteristics is then augmented with variables that measure the status of literacy in the district during the colonial period so that I am conditioning on the district's literacy starting point.[22]

One of the main constraints to the expansion of Christianity and the proselytizing activities of the missionaries was represented by Islam. In general, missionaries were less likely to locate their stations in areas with a strong Muslim presence (Richter [1908]). To take this into consideration, I also include the district's share of Muslim population in 1901.

The British control over India did not unfold homogeneously: some territories were administrated by the British as early as 1765 (through the *East India Company*), other areas were instead subjugated in a later period. The early exposure to the British control is likely to translate in an earlier adoption of the colonizer's legal system and institutional framework that may have attracted Protestant missions because of a higher cultural proximity to their countries of origin. To take this into account, using a map contained in the "Imperial Gazetteer of India" (see Figure 3), I create a set of dummy variables indicating if, as of 1837, the district was already under British control or if it was instead governed by a local ruler (distinguishing between Hindu and Muslim princes).

[Insert Figure 3 about here]

From the middle of the nineteenth century, the Government of India starts a great transportation infrastructure project aimed at developing a vast railroad network throughout the Indian sub-continent. Between 1853 and 1930, a total of 67,000 km of railroads are built, penetrating inland districts and bringing significant changes to the technology of trading in India. It is quite possible that railroads play a significant role in defining the missionaries' location decisions. *Ceteris paribus*, areas with a more developed railways network are more likely to be targeted by the early missionaries as more easily accessible. To take this into account, based on historical maps contained in the "Constable's Hand Atlas of India",[23] I construct a variable measuring the districts' diffusion of railways as of 1891.

[Insert Figure 4 about here]

At this point I have two distinguished data sets. On one hand I have data for the districts in 1901; on the other hand I have information for the districts in 2001. The next step consists of combining these two groups together. In order to do so I would need a conversion table to match the historical

[22] India was, essentially, an illiterate country in the nineteenth century. The census of 1901 reports that the literacy rate is about 4 percent for the whole population, 9 percent for the male population and 0.5 percent for the female population.

[23] Edited by John Bartholomew and Sons in 1893. The maps that show the status of the railways network in 1891 are contained in plate 18, 19 and 20.

and the modern districts. Unfortunately such conversion table does not exist. Luckily it is possible to build one. The "India Administrative Atlas, 1872-2001" keeps track of all the administrative changes occurred in India in the last 140 years. Using that source of information, I carefully match each district in 2001 with the corresponding territory in 1901. Unsurprisingly, in one century there have been several changes which, in some cases, prevent a one-to-one matching between historical and modern districts (on average a district in 1901 matches with about 1.7 districts in 2001). When this happens, the modern districts share, necessarily, the same historical information.

3.2 Contemporary and Geographic Data

I measure the districts' literacy status using variables from the 2001 Census of India.[24] According to the definition given in the census, an individual is considered literate if older than six and "able to read and write with understanding in any language".[25] Consistent with that definition, I use, as two of my outcome variables, the districts' total population and female population literacy rate as calculated by the census.[26] As previously discussed, the Protestant missionaries were among the few institutional actors to provide basic education to women. I therefore expect female literacy to be positively related to the exposure to historical missions. To test this hypothesis, I consider, as additional literacy outcome, an index ("GPI") that, consistent with the definition given by UNESCO, measures the degree of literacy gender parity in a society.[27] A larger GPI denotes a higher gender equality in terms of literacy.

The "Village Directory" of the census contains information regarding the diffusion, in each district, of several amenities. To account for differences in the degree of development across districts, I include variables that measure each district's share of villages with: access to electricity (for either domestic or agricultural use), medical facilities, a bus stop, a paved entry road. Similarly, the district's supply of education is approximated by the share of villages with a primary school, a secondary school, a college/university. Moreover, to account for the degree of urbanization, I also include the district's share of population living in urban areas.

Given its territorial extension, the Indian sub-continent is characterized by some degree of geographic and climatic heterogeneity. To take this into consideration, following the suggestion of, among others, Becker and Woessmann (2009) and Michalopoulos and Papaioannou (2013), I rely on an extensive set of geographic variables. To account for periphery I include the latitude and longitude (and their interaction) of the centroid of the district and its minimum distance (in kilometers) from one of the

[24] See http://www.censusindia.gov.in/2011-common/census_data_2001.html.

[25] Source: "Literacy in the Context of the Constitution" (2001), available at http://lawmin.nic.in/ncrwc/finalreport/v2b1-5.htm.

[26] The literacy rate is defined as $\frac{\text{Number of Literates}}{\text{Population Aged 7+}} \times 100$. See http://www.censusindia.gov.in/Metadata/Metada.htm [Last accessed: July 13th, 2013].

[27] GPI $= \frac{\text{Female Literacy Rate}}{\text{Male Literacy Rate}}$ (http://www.unescobkk.org/fileadmin/user_upload/aims/Pattaya_SCB_Jun05/Gender_Parity_Index_-_Michael_Koronkiewicz.pdf [Last accessed: July 13th, 2013]).

ten major cities of India.[28, 29] Motivated by the arguments of Rappaport and Sachs (2003) that coastal access is a fundamental determinant of economic outcomes, I include a dummy variable equal to one if the district has a coast and zero otherwise. Finally, to account for the potential effect of water streams and elevation on development (see, for example Nunn and Puga [2012]), using data from, respectively, the "Digital Chart of the World" and the "CGIAR Consortium for Spatial Information", I compute the total area covered by rivers per squared kilometer and the average altitude of each district.

3.3 Summary Statistics and Preliminary Evidence

Table 2 reports some descriptive statistics for all the variables used in the analysis.

The top panel focuses on the three outcome variables. On average 63 percent of a district's population is literate. As expected, this value drops significantly (to only 51 percent) when looking at the female population, suggesting the existence of a wide literacy gap between genders. This is confirmed by looking at the average value of the gender parity index, equal to 0.68 and well below the benchmark that defines literacy gender equality.

The second panel (labeled "Current Variables") shows that, on average, 80 percent of the villages have connection to electricity, 60 percent can be accessed by a paved entry road, 44 percent have at least a bus stop and 40 percent have a medical facility (of any kind) in their territory. Moreover, while the large majority of a district's villages have a primary school, less than one fourth have structures offering secondary education and only 1 percent have a college or university under their jurisdiction. Data also indicate that the distribution of these development variables is highly unequal and skewed, as suggested by the large values of the standard errors and the significant gaps between means and medians or between the twenty-fifth and the seventy-fifth percentiles.

Turning to the set of historical variables, it can be noticed how Indian districts, 60 percent of which were under direct British control during the early colonial period, are characterized by some heterogeneity in terms of their initial level of development as measured by the population density, the degree of urbanization and the share of non agricultural labor force in 1901. The bottom panel of the table shows that 283 districts (55 percent of my sample) were inhabited by Protestant missions as of 1908, with the number of missions ranging between one and ten.

[Insert Table 2 about here]

Figure 5 offers a visual inspection of the relationship between historical Protestant missions and current literacy outcomes. A map similar to the one presented in Figure 2 is reproduced with the difference that, now, the districts have different colors according to how they perform in terms of literacy. The

[28] I calculate district centroids using the "centroid" feature in ArcGIS.

[29] Based on data from 2001, the ten most populous cities in India are (in descending order): Mumbai, Delhi, Bangalore, Hyderabad, Ahmedabad, Chennai, Kolkata, Surat, Pune and Jaipur.

first panel, which considers the total population literacy rate, shows that there is an apparent clustering of historical missions in the districts characterized by higher levels of literacy (darker shade of blue), suggesting the existence of a positive association between historical missionary activity and the outcome variable. A similar conclusion can be drawn in relation to the female literacy rate (panel B) and the gender parity index (panel C).

[Insert Figure 5 about here]

4 Empirical Approach

4.1 Baseline Analysis

To formally examine the relationship between historical missions and literacy outcomes, I compare districts' literacy outcomes according to whether they were exposed to Protestant missions at the beginning of the twentieth century by running ordinary least squares (OLS) regressions of the form:

$$Y_{i,j}^{2001} = \alpha PM_{i,j}^{1908} + \mathbf{X_{i,j}}'\Gamma + \sum_{j=1}^{J} \theta_j d_j + \epsilon_{i,j} \tag{1}$$

where $Y_{i,j}^{2001}$ represents one of the current literacy outcomes in district i belonging to state j; $\mathbf{X_{i,j}}$ is a vector of exogenous controls; and $\epsilon_{i,j}$ is an unobserved error term. The variable $PM_{i,j}^{1908}$, the regressor of interest, is a binary variable equal to one if Protestant missions were operating in the district as of 1908, and zero otherwise. The parameter α measures the long-term relationship between historical Protestant missions and the current districts' literacy outcomes. In most specifications I include state dummies (d_j), so as to exploit only within-state variation.[30]

As always when considering OLS estimates, caution is necessary in interpreting these results. One may be concerned that they are only the consequence of a spurious correlation due to either reverse causality or omitted variables. Since missions were established more than a hundred years ago, they are clearly pre-determined with respect to the the outcome variable. Hence, reverse causality does not represent an issue in this context. Protestant missionaries, however, did not locate randomly across the Indian sub-continent. If there are omitted factors that affect the current literacy outcomes of the districts as well as the missionaries' location decisions, then it would not be possible to give a causal interpretation to the estimate of α in Equation 1. I address this issue in several different ways.

The data I use are at the district level, a district in India being an administrative division below

[30]The dependent variable (share of literates) is a proportion bounded between zero and one. There is some debate in the statistical and econometric literature regarding the most appropriate model to employ when the left-hand side variable is a proportion. By far the most common practice among researchers is to apply OLS. That is, for example, the choice adopted by Becker and Woessmann (2009). Others (see for example Papke and Wooldridge [1996]) propose to estimate models with fractional data using a generalized linear model (GLM) with a logit link and the binomial family. Johnson et al. (1995) provide several examples from different physical sciences in which the beta distribution is found to be a better fitting distribution for the proportional data. The use of either a GLM or a Beta-distribution based regression as alternatives to OLS produces equivalent results (available upon request).

state level. The immediate advantage of this choice is that, by exploiting regional variation within a country, I condition on a quite uniform legal and institutional framework. In addition, I also control for the historical political division of India, differentiating between areas that, during the nineteenth century, were directly under the British control and areas that were instead administratively autonomous (the so called "feudatory" or "princely" territories) to take into account the fact that the missionaries may have been drawn to areas culturally more similar to their countries of origin.

To exclude the possibility that the results are driven by geographic differences across the districts, I include a set of geographic controls. In most specifications I also introduce state fixed effects, using therefore only the variation across districts within each state. To the extent that there is unobserved regional heterogeneity (in terms of state-level institutions, for example), the state dummies capture most of its substance.

It may be possible that missionaries located themselves in districts that were intrinsically different (in terms of socio-economic characteristics) and that these differences perpetuated over decades and continue to exist today, too. To take this into account, I condition on the extensive set of historical controls that measure the level of socio-economic development of the districts at a time close to when the missionaries were settling in India.

4.2 Neighbor Districts Analysis

One may argue that the previous set of controls is not enough in order to account for all the factors that are correlated with both $Y_{i,j}^{2001}$ and $PM_{i,j}^{1908}$ in Equation 1. To mitigate such concern, I adopt two strategies.

First, following Banerjee and Iyer (2005) and Michalopoulos and Papaioannou (2013), I perform an analysis limited only to neighbor districts that differ in terms of their exposure to historical Protestant missions. This strategy is based on the idea that unobserved characteristics that may not be captured by the extensive set of variables mentioned before should be similar for adjacent districts. Consequently, differences in outcomes between neighbor districts are more likely to be due to the exposure to historical Protestant missions.

I consider only those districts that share a common border and happen to differ in the sense that one of them was inhabited by Protestant missions as of 1908, while the other one was not. For the sake of clarity, the figure below, gives a stylized representation of this strategy. The rectangle identifies a representative state. The squares inside it define the different districts. Three districts are inhabited by historical Protestant missions. In the previous analyses, all the districts (squares in this example) entered the estimation sample. Now, instead, the estimation is based only on the shaded districts, divided in three different groups. Following Michalopoulos and Papaioannou (2013) I first

verify whether there is a systematic correlation between the exposure to historical missions and several characteristics within groups of neighbor districts in the same state. To this aim, I run least squares specifications of the following form:

$$Z_{i(g),j} = \alpha PM^{1908}_{i(g),j} + \sum_{j=1}^{J}\sum_{g=1}^{G}\theta_{g,j}d_{g,j} + \eta_{i(g),j} \tag{2}$$

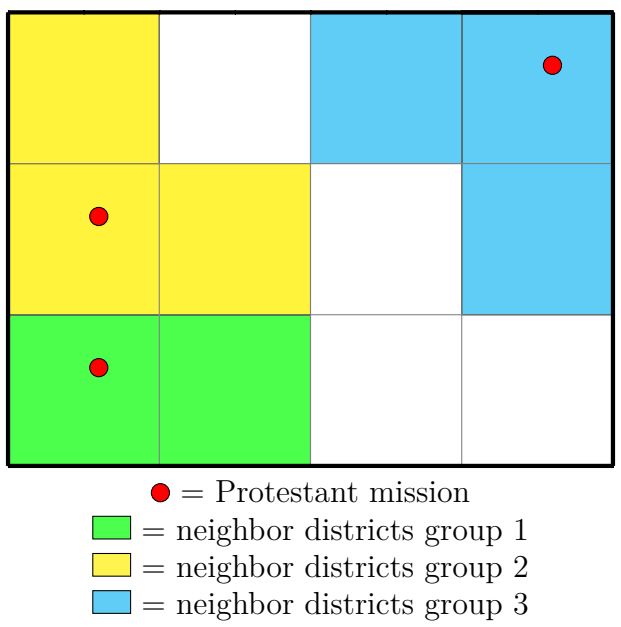
where the subscript $i(g), j$ denotes the district i, part of the adjacent-districts group g, in state j; the vector Z contains geographic (distance to major city, coastal dummy, rivers, mean altitude) and historical (urbanization, population density, share of non-agricultural workforce, share of muslims, infirmities and literacy rates in 1901) variables; PM is the Protestant indicator variable as defined before; and the terms $d_{g,j}$ are neighbor-districts-groups specific dummies.

I then examine whether the exposure to historical Protestant missions leads to differences in current literacy outcomes within contiguous districts in the same state by estimating regressions of the form:

$$Y^{2001}_{i(g),j} = \alpha PM^{1908}_{i(g),j} + \mathbf{X_{i,j}}'\Gamma + \sum_{j=1}^{J}\sum_{g=1}^{G}\theta_{g,j}d_{g,j} + \epsilon_{i(g),j} \tag{3}$$

Given the inclusion of neighbor-districts-groups specific fixed effects $d_{g,j}$, the coefficient α captures whether differences in historical Protestant density translates into difference in contemporary literacy outcomes across adjacent districts belonging to group g, in state j.

Neighbor Districts



4.3 Using Catholic Missions as Control Group

The second strategy consists of using historical Catholic missions as control group. Data about the location of Catholic missions in India during the colonial period come from the "Atlas Hierarchicus". This Atlas, which represents an official document of the Vatican, was published in 1913 and provides detailed maps with the worldwide location of Catholic missions as of 1911. Figure 6 shows a section of one of the charts contained in the Atlas. Using the same geo-referencing techniques in a GIS environment explained before, I combine that historical information with the districts' administrative boundaries in 2001.

[Insert Figure 6 about here]

Inspired by Donaldson (2010) I explore the plausibility of concerns about potential bias due to endogenous Protestant placement by estimating the effect of Catholic missions. Following the "Roman Catholic Relief Act" of 1829, the British Crown maintained a neutral position toward the activities of Christian missionaries in its colonies. Both Protestant and Catholic were free to operate in India without any particular restriction. It is therefore plausible to assume that missionaries belonging to either denominations were influenced by the same factors in deciding where to locate. The two groups, however, differed with respect to their attitudes toward the promotion of universal education. While Protestants had a keen interest in the promotion of mass literacy and were especially committed to educate non-elites and socially disadvantaged groups (such as women and low castes members), Catholic missionaries, at least prior to the Second Vatican Council in 1965, predominantly invested in schools dedicated to the formation of priests and elites only.[31] Therefore, I do not expect the historical Catholic influence to have a significant association with mass literacy and I can use Catholic missions as a "placebo" group. If location decisions were driven by unobserved determinants of current literacy outcomes, then Catholic missions would exhibit spurious effects (relative to the excluded category, districts in which there were no missions at all) on literacy outcomes in OLS regressions of the form:

$$Y_{i,j}^{2001} = \alpha PM_{i,j}^{1908} + \beta CM_{i,j}^{1911} + \mathbf{X_{i,j}}'\Gamma + \sum_{j=1}^{J} \theta_j d_j + \epsilon_{i,j} \tag{4}$$

where $CM_{i,j}^{1911}$ is a binary variable equal to one if a Catholic mission was active in district i as of 1911, and zero otherwise. Results from this regression can provide an additional indication of whether, and to what extent, the estimate of α in Equation 1 is driven by omitted districts' characteristics.

[31] Prior to 1965, the Catholic Church did not encourage lay people to read and interpret the Bible. Catholic missionaries seldom translated the Bible into local languages and put less emphasis on mass education.

5 Results

5.1 Baseline Estimates

Table 3 reports cross-sectional least squares specifications that associate the district's current literacy outcomes (each panel of the table corresponds to one of the three outcomes) with the presence of Protestant missions as of 1908. The reported standard errors are robust to heteroskedasticity and clustered at state level to control for the potential correlation of disturbances across districts that belong to the same state. Column [1] shows the unconditional estimates. Consistent with what suggested by Figure 5, historical "Protestant" districts perform better in terms of contemporary literacy outcomes relative to their "non-Protestant" counterparts.

Column [2] includes state fixed effects, to exploit only within state, cross-districts variation. As expected, this specification produces more conservative estimates. For all the three literacy outcomes, the introduction of state fixed effects reduces the size of the coefficient on the regressor of interest as the regional heterogeneity is now captured by the state dummies; yet the estimates maintain significance at the 99 percent confidence level. The estimated coefficients suggest that historical "Protestant" districts have a population literacy rate, a female literacy rate and a gender parity index which is, respectively, 3.3, 4.1 and 3 percentage points larger than those of "non-Protestant" districts.

Geography represents an important fundamental of socio-economic development (Gallup *et al.* [1998]). While geographic differences may not be as dramatic in India as they are in a cross-country analysis, in column [3] I test the sensitivity of the previous findings against adding a set of geographic controls: the coordinates of the centroid of the district, its minimum distance from a major city, an indicator variable for coastal districts, a variable measuring the extension of the surface under water and the average altitude of the district to account for the potential positive effects on economic development. Conditioning on these geographic measures has little effect on the main results.[32]

Finally, in column [4] I include historical information to control for factors that may have affected the location decisions of the early missionaries. Variables that measure the degree of historical development of the district (population density, degree of urbanization and the share of non-agricultural workforce in 1901, density of railways in 1891) are all positively related to the current status of literacy but their estimated coefficients are only mildly significant.[33] Regions in India did not fall under the British control simultaneously. To account for the fact that some areas have been colonized earlier than others and have therefore been in contact with British institutions for a longer time, I add a set of dummy variables indicating if the districts belong to an area that, as of 1837, was under direct British

[32] The coastal dummy enters the models with a positive and statistically significant coefficient. Average altitude and the minimum distance from a major city enter with negative, although not significant, estimates. The coefficient on the variable measuring the extension of water streams in the district is positive and, in some specifications, significant. Results available upon request.

[33] The variables that enter the regression with a strongly significant coefficient are: degree of urbanization (positive), share of muslim population (negative), historical literacy outcome (positive).

administration or under the control of a local (Hindu or Muslim) prince. I also include measures of literacy outcomes during the colonial period so that the resulting estimate is conditional to the initial literacy status of the districts. Even when this extensive set of historical controls is added to the model, the coefficient on Protestant density maintains its positive association with the dependent variable and its statistical significance.

The point estimate implies that, on average, having had Protestant missions during the early colonial period is associated with a 2.5 percentage points increase of the total population literacy rate. As hypothesized, given the central involvement of the Protestant missionaries in the matter, historical missions appear to be important also for the advancement of female literacy: the estimated coefficient implies that historical "Protestant" districts have a current female literacy rate which is 3 percentage points higher that that of areas that have not seen the presence of Protestant missionaries. The findings also suggest that historical missions benefit women relatively more than men: the gender parity index is about 2 points higher in "Protestant" districts relative to areas with no missions.

[Insert Table 3 about here]

Although the previous results suggest that being exposed to historical Protestant missions seem to be consistently associated with better contemporary literacy outcomes, one may still be worried that the rich set of conditioning variables included in the previous specifications does not fully capture all the factors correlated with the location decisions made by the early missionaries. If these unobservables are also correlated with the outcome variables (becoming, therefore, a component of the error term $\epsilon_{i,j}$ in Equation 1), then the interpretation of the previous findings has to be tempered by the possibility that these may be driven by omitted variables. In that case, my results would uncover only a simple correlation rather than a causal effect. To be interpreted as causal, unbiased estimates, the exposure to Protestant missions (i.e. $PM_{i,j}^{1908} = \{0, 1\}$) has to be conditionally random. To examine this issue formally, I adopt the Altonji et al. (2005) method of using the degree of selection on observables as a guide for selection on unobservables. This method relies on the comparison between a regression that includes potentially confounding factors, and one that does not. Defining α^F the estimate from the "full" regression and α^R the estimate from the "restricted" regression, the Altonji-Elder-Taber (AET) ratio is equal to $|\alpha^F/(\alpha^R - \alpha^F)|$. When selection on observables does not significantly affect the estimate, the denominator is small, increasing the ratio. The smaller is the impact of observables on the estimate, the larger the impact of unobservables has to be in order to reduce the effect of historical Protestant exposure. Nunn and Wantchekon (2011) suggest that, as a rule of thumb, any ratio above one is acceptable since it indicates that selection on unobservables must be large than selection on observables in order to explain the effect. Table 4 shows the AET ratios. The "full" specification corresponds to column [4] of Table 3, while the "restricted" specification excludes both geographic and historical controls (column[2] of Table 3). In all cases, the ratio is well above one, ranging between

2.3 and 3.1. Therefore, to attribute the entire OLS estimate to endogenous omitted characteristics, selection on unobservables would have to be at least twice as big as selection on observables. I interpret this result as suggestive of the fact that it is unlikely for the estimated effect of historical Protestant exposure to be fully driven by unobservables.

In the following sections I examine this issue further by undertaking two alternative strategies.

5.2 Neighbor Districts Estimates

The first strategy (suggested in Banerjee and Iyer [2005] and employed also in Michalopoulos and Papaioannou [2013]) consists of exploiting within-state, within neighbor-districts variation in Protestant exposure and literacy rates. The underlying idea is that unobservable characteristics that may not be captured by the extensive set of controls introduced in the previous specifications should be, in fact, similar for contiguous districts. Therefore, differences in literacy outcomes between neighbor districts are likely to be due to differences in the historical exposure to Protestantism.

I consider groups of contiguous districts where one district saw the presence of Protestant missions during the nineteenth century while the neighboring one did not. Table 5 shows that by focusing on these groups of districts the role of local geographic and historical factors is neutralized as they are never significantly correlated with the exposure to historical missions.[34]

As shown in Table 6, even when focusing on groups of very similar districts, there is still clear evidence of a beneficial role of Protestant missions. The estimates confirm that within states and within groups of contiguous districts, literacy outcomes are significantly better in "Protestant" districts than in the "non-Protestant" counterparts. I interpret these findings as indicative of the fact that the long-term positive relationship between missions and literacy is unlikely to be driven unobserved districts' attributes.

[Insert Table 5 about here]

[Insert Table 6 about here]

5.3 Using Catholic Missions as Control Group

As mentioned previously, the British Crown adopts a neutral policy toward missionary activities in its colonies, so that missionaries of every Christian denomination are free to operate in India without any specific restriction.[35] The data confirm that it is common for a district to have both Catholic

[34] I also verify, via a t-test, that within the sample of neighbor districts "Protestant" and "non-Protestant" areas display no heterogeneity over observable characteristics. Results are contained in Table A.1.

[35] This differs from the policy adopted by other colonial powers. Spanish, Portuguese and Italian colonies have a more explicit bias toward Catholic missionaries. This takes a form of a number of regulations favoring the Catholic Church and limiting the influence and actions of Protestant missionaries. The Portuguese, for example, allowed Protestants to enter Angola and Mozambique, but banned Protestant missions from being located near Catholic missions (generally about thirty miles).

and Protestant missions: in the majority of the cases (75 percent) Catholic missionaries operate in a district populated also by Protestant groups. It is therefore plausible to assume that similar factors determined the location decisions of the two denominational groups. Based on this assumption, I check the sensitivity of my results to including a Catholic dummy (equal to one if a Catholic mission was in the district as of 1910, and zero otherwise) to the set of regressors in Equation 1. The underlying idea is that if location decisions are driven by unobserved determinants of contemporary literacy outcomes (common to both Catholic and Protestant missionaries), then Catholic missions would exhibit spurious effects (relative to the excluded category, districts in which there are no missions at all) on the outcome variables in OLS regressions with state fixed effects. However, as showed in Table 7, for all the three literacy outcomes, the estimated coefficients on the Catholic dummy are never statistically different from zero or of the same order of magnitude as those of the Protestant dummy. Crucially, the inclusion of the Catholic dummy never significantly affects the magnitude of the Protestant coefficient. This result casts additional doubt on the extent to which Protestant missionaries were selecting districts for their missions on the basis of correlation with the error term in Equation 1.

[Insert Table 7 about here]

Taken together, the analyses based on either neighbor districts or the use of Catholic missions as control group serve to confirm that the baseline results are not primarily caused by some unobserved districts' characteristics.

5.4 Protestant Missions and the Literacy of Low Castes

One of the distinguishing features of the Protestant missionary movement in India is its focus on the poor and marginalized members of society. Protestant and Catholic missionaries show great differences regarding their attitudes toward the lower castes. Catholics generally accept the caste system and use it to build a strong position in the Indian society. In their view, Christianity is seen neither threatening nor undermining the caste system but rather working within it and accommodating western social standards to the norm of castes:

> *"By becoming a Christian, one does not renounce his caste, nobility or usage. The idea that Christianity interfered with them has been impressed upon people by the devil, and is the great obstacle to Christianity"*
>
> [letter from Robert de Nobili, S. J. as reported in Forrester (1980)]

> *"There is no anarchical wish to tear up the rails along which the train of Indian society has run for over a thousand years"*
>
> [Joseph C. Houpert, S. J. (1937), *A South-Indian Mission: The Madura Catholic Mission from 1535-1935*, page 257]

A clear example of this attitude is represented by the case of the Jesuit mission in Madurai (Tamil Nadu) centered around different church buildings for high-castes and low-castes members. The *dalits* (so called "untouchables") were in any case kept outside the church.

On the other hand, Protestant missionaries view the caste system as an obstacle not only to conversions but to enlightenment and progress as well. By 1850, the various Protestant denominations agree that caste within the Church represents an "unmitigated evil" and that there is a specific Christian responsibility to undermine the caste system as such.[36] Education, in particular, is used as an invaluable weapon in the fight against castes. There is the belief that Christian schools cannot recognize or tolerate caste observances within their walls. The Protestant stance is that education must be accessible to all:

> *"In all the new institutions the important principle has been established of admitting boys of every caste without distinction* [...]*Christian, Mohammedan and Hindu boys, of every shade of colour and variety of descent, may be seen standing side by side in the same class, engaged in the common pursuit of English literature, contending for the same honours, and forced to acknowledge the existence of superior merit in their comrades of the lowest as well as the highest cast."*
>
> [C. E. Trevelyan (1838), *On the Education of People in India*, page 19]

In Protestant schools the pupils were distributed into classes irrespectively of age and caste, giving therefore equal educational opportunities to members of the lowest social groups. This is likely to be responsible for the fact that over that period "Protestant" *dalits* generally have higher literacy rates than their "non-Protestant" peers (Massey [2007]). It is therefore interesting to see if the beneficial effect of Protestantism on the literacy of the lowest classes of Indian society has persisted over decades until today. To this aim I consider, as my new dependent variable, the literacy rate of individuals that belong to the so called Scheduled Castes and I repeat the same analysis presented in the previous section.

[Insert Table 8 about here]

Table 8 shows that, across the different specifications, there is consistent evidence of a positive effect of Protestantism on the literacy of historically disadvantaged castes. In "Protestant" districts, scheduled castes population exhibits literacy rates that are between two and five percentage points higher relative to those in "non Protestant" areas.

[36] Many Protestant missions require new converts to prove their rejection of the caste system. The American Madurai Mission, for example, insisted that new members of the community took part of so-called "love-feasts" where they ate with missionaries and Christians from other castes food usually prepared by a low-caste cook. For additional details about the relationship between Christian missionaries and the caste system in India, please refer to Gladstone (1984) and Bugge (1998).

6 Robustness Checks

6.1 Outliers

Table A.2 in the appendix reports estimates of Equation 1 when excluding observations that fall at either the upper or lower tail of the dependent variables' distribution. Across the three literacy outcomes, I find that once the extreme values are excluded the coefficient on the Protestant dummy increases. I also repeat the estimation dropping the districts that, previous to the British domination, had been under the control of the Portuguese Crown (states of Goa, Kerala and Tamil Nadu). This adjustment does not affect the results in any significant way.

[Insert Table A.2 about here]

6.2 Alternative Measures of Protestant Influence

Using a binary indicator as measure of historical Protestant influence has the advantage of offering a relatively easy interpretation: the coefficient α in Equation 1 (once one makes sure that endogeneity issues are not driving its estimate) can be read as the treatment effect of being exposed to Protestant missions. The disadvantage is that, by relying on a binary measure, one completely discards the variation in the number of historical missions in the districts. As shown in Figure 7, within the sample of "Protestant" districts, the number of missions varies between one and ten and about 45 percent of the districts have at least two missions.

[Insert Figure 7 about here]

It may be that not only being exposed to the Protestant presence, but also the intensity of such exposure, plays a relevant role in determining current literacy outcomes. To investigate this point, in what follows I present results from the estimation of Equation 1 when the variable $M_{i,j}^{1908}$ is a continuous measure of Protestant influence.

6.2.1 Historical Protestant Density

To exploit the degree of exposure to Protestantism, I first use a variable ("Protestant Density") equal to the number of historical missions per squared mile in each district. The average district is inhabited by about 1.2 historical missions, translating in a "Protestant density" of 0.07 missions every 100 squared miles. Table A.3 reports the results from the estimation of Equation 1 when this continuous regressor is used. The first four columns refer to the full sample of districts, while the last column limits the analysis only to groups of neighbor districts as defined in section 4.2. Across the different specifications and samples, there is evidence of a positive association between the degree of exposure to historical

Protestant missions and current literacy outcomes. The table also reports the marginal effect due to a one standard deviation increase of Protestant density. The most conservative estimates suggest that one additional Protestant mission in the average district is associated with a 0.6 percentage points increase of the total population literacy rate; a 0.7 percentage points increase of the female literacy rate; a 0.3 points improvement of the district's gender parity index; and a 0.8 percentage points increase of the literacy rate of scheduled castes population.[37]

[Insert Table A.3 about here]

In some cases Protestant missions are located in the proximity of the border between two districts. Assume, for example that district A contains a mission near the border with district B. Then, technically, only district A would be labeled as having an historical mission in its territory. However it is plausible to think that part of the effect (if any) of the mission extends also to district B, given its vicinity. To take cases like this into account, around the borders of each district I draw a buffer with a varying radius and consider all the missions falling within these "buffered" borders as belonging to the district itself. Table A.4 shows that, irrespective of the size of the buffer, the results are qualitatively similar to the previous analysis. The magnitude of the estimated coefficients decreases with the length of the buffer's radius, because of the attenuation bias introduced by the measurement error in the density variable.

[Insert Table A.4 about here]

6.2.2 Average Minimum Distance

The second continuous measure of Protestant influence is represented by the district's average minimum distance from a mission. Building this measure requires two steps. First, within each of the 515 districts, I generate 10,000 randomly located points. Second, I calculate the distance (in kilometers) of each of these points from the nearest Protestant (and Catholic) mission so that I can compute the district's average minimum distance to Protestant (and Catholic) a mission.[38] Table A.5 reports the results of the estimation of Equation 1 using this new variable as measure of historical Protestant influence. In line with the previous findings, the estimated coefficients suggest that the more distant from a Protestant mission a district is, the worse its current literacy outcomes. The coefficients in columns [1]-[5] imply that, for the average district, a 10 kilometer increase of its average minimum distance from a Protestant mission determines a decrease of about 1 percentage points of the total population and the female literacy rate, of 1.3 percentage points of the literacy rate of low castes, and of 0.4 points of the literacy gender parity index. On the other hand, variations in the district's average minimum

[37] Lewbel (2012) proposes an estimation method to identify parameters in regressions with potentially endogenous variables when traditional identifying information, such as external instruments or repeated observations, is not available. When I implement that methodology I obtain estimated coefficients that are close to those shown in column [4] of Table A.3 (results available upon request).

[38] These two steps have been implemented using the "Random Points Generator" and the "Proximity Analysis" packages in ArcGIS.

distance from a historical Catholic mission do not have any effect on the dependent variables and the corresponding estimated coefficients are never of the same order of magnitude of those associated with Protestant missions. The results are also confirmed if the analysis is limited only to districts that had either a Protestant or a Catholic mission and to districts that had both groups of missionaries.

[Insert Table A.5 about here]

7 Potential Transmission Channels

The results presented in Section 5 support the existence of a persistent positive treatment effect of Protestant missions, established in India more than a century ago, on current literacy outcomes. Through which channels does this effect operate? What follows will discuss possible answers to that question.

A first potential channel is represented by the positive role that missionaries had in increasing the supply of education in the areas where they were operating. As previously discussed, it was common among missionaries to establish educational facilities (from simple reading classes to boarding schools and colleges to "Zenana" schools for women). Some of these institutions preserved over time so that Protestant districts may currently display better literacy outcomes simply because they are characterized by a higher supply of schools. To test this point I add to the set of controls the district's share of villages with, respectively, primary schools, secondary schools and universities/colleges.

Another possibility is that missions were instrumental for the future development of the districts. For example, by maintaining contact with their countries of origin and constant access to supplies from Europe, the presence of missionaries may have improved the districts' openness to trade fostering, consequently, their economic development. To assess the importance of this channel I include a set of variables that measure the degree of development of the district (degree of urbanization; share of district's villages with access to electricity, medical facilities, a bus stop, a paved entry road).

One may also think that, by targeting only those who converted to Christianity, missionaries were able to influence only a restricted (and relatively small) group of the population. If that were the case, I should find that my results depend on the size of the district's Christian community. This is unlikely to have happened in the Indian context as, since the early development of their settlements, missionaries did not limit their educational offer only to those who converted to Christianity. On the contrary, conversions represented a potential harm for missions' schools as they prompted Hindu families to withdraw their children from school. To ensure the survival and the success of their schools, Protestants needed to attract people from other religious backgrounds and, to this aim, they necessarily had to put aside their mission to evangelize and declare the Christian faith.[39] For this reason I do not

[39]Three anecdotal stories can be mentioned to support this point. Wilson (1928) reports of a man in south India trying to convince parents to send their daughters to a mission school telling them: "*Why do you not send your daughters to this school? Are you afraid of the Christian*

expect my results to be significantly affected by the size of the district's Christian community.

Table 9 reports the results obtained when these additional sets of variables are added to the model. The district's supply of education correlates positively with the dependent variables. Among the three types of schools, secondary schools and colleges/universities seem to be particularly beneficial to current literacy outcomes. The higher the degree of urbanization of the district, the better the literacy outcomes. Among the variables that indicate the diffusion of several "amenities" in the district, access to electricity is the only one to be significantly correlated with higher levels of literacy. Having a larger share of Christian population does not seem to have a significant role in affecting the literacy of the district. Notice that when these groups of variables are included, the effect of Protestant missions is attenuated (the estimated coefficients reduce their magnitudes by around one-third) but is still statistical significant at the 95 percent confidence level. Applying recent work by Gelbach (2009), the table also reports the degree to which the three different groups of covariates reduce the size of the effect of Protestant missions on literacy. The description of the methodology is provided in Appendix B. Almost all of the reduction in the estimated coefficients appear to be due to the inclusion of the development controls.

[Insert Table 9 about here]

Taken together, these results indicate that part of the effect of historical missions certainly works via the impact that the early Protestant missionaries had on the subsequent level of socio-economic development of the districts. However, a significant portion of this relationship remains unexplained and operates through some other channel. One possible explanation is that part of the positive long-term effect of historical missions is demand driven, that is due to a persistent change in people's attitudes toward education. By preaching the importance of education and the acquisition of literacy skills, the first missionaries may have changed people's views about the relevance of being educated. It is likely that these individuals transmitted such attitudes and beliefs to their children so that the descendants of those in contact with the early missionaries value education more and are therefore more literate today. This is in line with a process of cultural transmission according to which beliefs and cultural traits that originate in the past may persist over several decades.[40] Moreover, this change in attitudes is not limited to the Christian community, suggesting that it is the result of a broader cultural change rather than just a direct religious effect.

teaching that is given here? Do you think your daughter will become a Christian? If so, be comforted. In thirty years experience, I have never known that to happen". Bellenoit (2007b) mentions the story of a mission school outside Lucknow that lost its students and was forced to close when rumors circulated that teachers were trying to convert students. Finally, according to a report of the London Missionary Society, in the spring of 1891 an open rebellion erupted in Almora (Uttarakhand) when a Brahmin student of Ramsay College converted to Christianity. The headmaster, Edward Oakley, admitted that conversions would *"empty our classrooms"* (cited in Bellenoit [2007b]).

[40]The analysis of the intergenerational transmission of values, preferences and beliefs has been at the center of several works on anthropology and sociology since the early 1980s (see in particular Cavalli-Sforza and Feldman [1981], Boyd and Richerson [1985]). The first to translate this process of cultural transmission in economic terms are Bisin and Verdier (2000 and 2001).

8 Conclusion

One of the pillars of the Protestant theology, the principle of "Sola Scriptura" implies that interpretations of the Scripture do not have the same authority of the Scripture itself. In order to understand God's word, a good Christian only needs to read the Bible which is self-authenticating. Quite obviously, to read the Bible, one has first to be able to read. Therefore, to ensure the fulfillment of the principle of "sola Scriptura", and to give everyone the capability of reading and understanding the Bible, Protestants stressed the importance of education and literacy among all the groups of population. Accordingly, the hypothesis of this paper is that Protestant missionary activities that occurred between the end of the nineteenth and the beginning of the twentieth century represent an important factor to explain current (geographical and gender) disparities in literacy outcomes in India.

Using an original data set that combines both modern and historical sources of information, I consistently find evidence of a persistent positive relationship between historical Protestant missions and current literacy outcomes. The exposure to Protestant missions (measured at either the extensive or intensive margin) is associated with significant improvements of literacy outcomes, particularly for the most disadvantaged groups of society (women and low castes).

The paper also verifies that these results are not the product of spurious correlation due to unobserved district characteristics that affect both current literacy as well as the early missionaries' location decisions. An analysis limited to groups of geographically adjacent districts that differ only in terms of their exposure to Protestantism reassures that omitted variables are unlikely to be responsible for the uncovered positive relationship. A placebo experiment using information about the location of Catholic missions around the same period of time lead to similar conclusions.

The long-term positive effect of missions on literacy can operate through different channels. I see that part of this relationship is explained by the positive role that missions had in laying the foundations to the future development of the districts. A significant portion of this relationship, however, remains unexplained and operates through some other mechanism. It is possible that part of the positive long-term effect of missions on literacy is demand driven: by preaching the importance of being literate the first missionaries changed people's views toward education and these new attitudes persisted over decades through a process of intergenerational cultural transmission.

This paper contributes to illustrate the mechanisms through which disparities arise and persist over time, in line with a recent and exciting literature looking at history and historical institutions as important determinants of contemporary development.

Bibliography

Acemoglu, D. and J. Angrist (2000). How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. NBER Macroeconomics Annual, National Bureau of Economic Research. [2]

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review 91*(5), 1369–1401. [2]

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics 117*(4), 1231–1294.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy 113*(1), 151–184. [3], [20]

Bainton, R. (1950). *Here I Stand. Life of Martin Luther*. Nashville, U.S.A.: Abingdon Press. [2]

Banerjee, A. and L. Iyer (2005). History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *The American Economic Review 95*(4), 1190–1213. [2], [3], [5], [16], [21]

Barro, R. J. and R. M. McCleary (2003). Religion and Economic Growth Across Countries. *American Sociological Review 68*(5), 760–781. [2]

Barro, R. J. and R. M. McCleary (2005). Which Countries Have State Religions? *The Quarterly Journal of Economics 120*(4), 1331–1370. [2]

Basten, C. and F. Betz (2013). Beyond Work Ethic: Religion, Individual, and Political Preferences. *American Economic Journal: Economic Policy 5*(3), 67–91. [4]

Beach, H. P. (1908). *India and Christian Oppoprtunity*. New York, U.S.A.: Student Volunteer Movement for Foreign Missions. [6]

Becker, S. and L. Woessmann (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *The Quarterly Journal of Economics 124*(2), 531–596. [4], [13], [15]

Bellenoit, H. J. A. (2007a). *Missionary Education and Empire in Late Colonial India, 1860-1920*. London, U.K.: Pickering & Chatto. [6]

Bellenoit, H. J. A. (2007b). Missionary Education, Religion and Knowledge in India, c. 1880-1915. *Modern Asian Studies 41*(2), 369–394. [27]

Berinsky, A. J. and G. S. Lenz (2011). Education and Political Participation: Exploring the Causal Link. *Political Behavior 33*(3), 357–373. [2]

Bisin, A. and T. Verdier (2000). Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits. *The Quarterly Journal of Economics 115*(3), 955–988. [4], [27]

Bisin, A. and T. Verdier (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory 97*(2), 298–319. [4], [27]

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *The American Economic Review 95*(1), 437–449. [4]

Boyd, R. and P. J. Richerson (1985). *Culture and the Evolutionary Process*. Chicago, U.S.A.: University of Chicago Press. [27]

Bugge, H. (1998, July-September). Christianity and Caste in XIXth Century South India. The Different Social Policies of British and Non-British Christian Missions. *Archives des Sciences Sociales des Religions 103*, 87–97. [23]

Cantoni, D. (2010). The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands. Working paper, Universitat Pompeu Fabra. [4]

Card, D. (1999). The Causal Effect of Education on Earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A. North Holland. [2], [9]

Cavalli Sforza, L. L. and M. W. Feldman (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, U.S.A.: Princeton University Press. [27]

Chevalier, A. (2004). Parental Education and Child's Education: A Natural Experiment. Working Paper 1153, IZA. [4]

Currie, J. and E. Moretti (2003). Mother's Education and the Intergenerational Transmission of Human capital: Evidence from College Openings. *The Quarterly Journal of Economics 118*(4), 1495–1532. [9]

Cutler, D. and A. Lleras-Muney (2006). Education and Health: Evaluating Theories and Evidence. NBER Workin Paper 12352, National Bureau of Economic Research. [2]

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica 78*(6), 1863–1906. [2], [5]

Diamond, J. M. (1997). *Guns, Germs and Steel: The Fate of Human Societies*. New York, U.S.A.: W. W. Norton and Co.

Dollar, D. and R. Gatti (1999). Gender Inequality, Income, and Growth: Are Good Times Good for Women? Gender and Development Working Paper Series 1, The World Bank. [9]

Donaldson, D. (2010). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. Working Paper 16487, National Bureau of Economic Research. [18]

Engerman, S. L. and K. L. Sokoloff (2002). Factor Endowments, Inequality, and Paths of Development Among New World Economies. Working Paper 9259, National Bureau of Economic Research. [2]

Forrester, D. B. (1980). *Caste and Christianity: Attitudes and Policies on Caste of Anglo-Saxon Protestant Missions in India*. London, U.K.: Curzon Press. [22]

Gallego, F. A. and R. Woodberry (2010). Christian Missionaries and Education in Former African Colonies: How Competition Mattered. *Journal of African Economies 19*(3), 294–329. [4]

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1998). Geography and Economic Development. Working Paper 6849, National Bureau of Economic Research. [19]

Gelbach, J. B. (2009). When Do Covariates Matter? And Which Ones, And How Much? Working paper 09-07, University of Arizona. [27], [47], [53]

Gennaioli, N., R. La Porta, F. Lopez de Silanes, and A. Shleifer (2012). Human Capital and Regional Development. *The Quarterly Journal of Economics 128*(1), 105–164. [2]

Gladstone, J. W. (1984). *Protestant Christianity and People's Movements in Kerala*. Trivandrum, India: Seminary Publications. [23]

Hanushek, E. A. and L. Woessmann (2012). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth 17*(4), 267–321. [2]

Ioannides, Y. M. (2003). Empirical Nonlinearities and Neighborhood Effects in the Intergenerational Transmission of Human Capital. *Applied Economic Letters 10*, 535–539. [4]

Jenkins, R. and E. Barr (2006). Social Exclusion of Scheduled Caste Children from Primary Education in India. http://www.unicef.org/files/Social_Exclusion_of_Scheduled_Caste_Children_from_Primary_Education_in_India.pdf. Last accessed: May 18th, 2013. [8]

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*. New York, U.S.A.: John Wiley and Sons, Inc. [15]

Kam, C. D. and C. L. Palmer (2008). Reconsidering the Effects of Education on Political Participation. *The Journal of Politics 70*(3), 612–631. [2]

Kaye, J. W. (1859). *Christianity in India: an Historical Narrative*. London, U.K.: Smith, Elder & Co. [5]

Kremer, M. (1997). How Much Does Sorting Increase Inequality? *The Quarterly Journal of Economics 112*(1), 115–139. [4]

Lewbel, A. (2012). Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business and Economic Statistics 30*(1), 67–80. [25]

Mantovanelli, F. (2012). Christian Missions, HIV and Sexual Behavior in Sub-Saharan Africa. mimeo. [10]

Massey, S. (2007). *Christian Missionaries in India*. Delhi, India: Sumit Enterprises. [23]

Mathew, A. (1988). *Christian Missions Education and Nationalism: from Dominance to Compromise, 1870-1930*. Shakti Nagar, Delhi, India: Anamika Prakashan. [6]

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial Ethnic Intitutions and Contemporary African Development. *Econometrica 81*(1), 113–152. [3], [13], [16], [21]

Moffett, S. H. (2007). *A History of Christianity in Asia Vol. 2: 1500-1900*. New York, U.S.A.: Orbis Books. [6]

Noel, B. W. (1837). *Duties of Christians Towards the Female Children of India and the East*. London, U.K.: Edward Suter. [7]

Nunn, N. (2009). The Importance of History for Economic Development. *Annual Review of Economics 1*(1), 65–92. [2]

Nunn, N. (2011). Christians in Colonial Africa. Unpublished manuscript. [5]

Nunn, N. and D. Puga (2012). Ruggedness: the Blessing of Bad Geography in Africa. *Review of Economics and Statistics 94*(1), 20–36. [14]

Nunn, N. and L. Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa. *The American Economic Review 101*(7), 3221–3252. [2], [20]

Oreopoulos, P. (2003). Do Dropouts Drop Out Too Soon? International Evidence From Changes in School-Leaving Laws. Working Paper 10155, National Bureau of Economic Research. [4]

Papke, L. E. and J. M. Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics 11*(6), 619–632. [15]

Psacharopoulos, G. and H. A. Patrinos (2004). Returns to Investment in Education: A Further Update. *Education Economics 12*, 111–134. [9]

Rappaport, J. and J. D. Sachs (2003). The United States as a Coastal Nation. *Journal of Economic Growth 8*(1), 5–46. [14]

Richter, J. (1908). *A History of Missions in India*. New York, U.S.A.: Fleming H. Revell Company. [12]

Sawage, D. W. (1997). Missionaries and the Development of a Colonial Ideology of Female Education in India. *Gender & History 9*(2), 201–221. [7]

UNESCO (2010). Report 2010: Reaching the Marginalized. http://www.unesco.org/new/en/education/themes/leading-the-international-agenda/efareport/reports/2010-marginalization/. Last accessed: May 18th, 2013. [8]

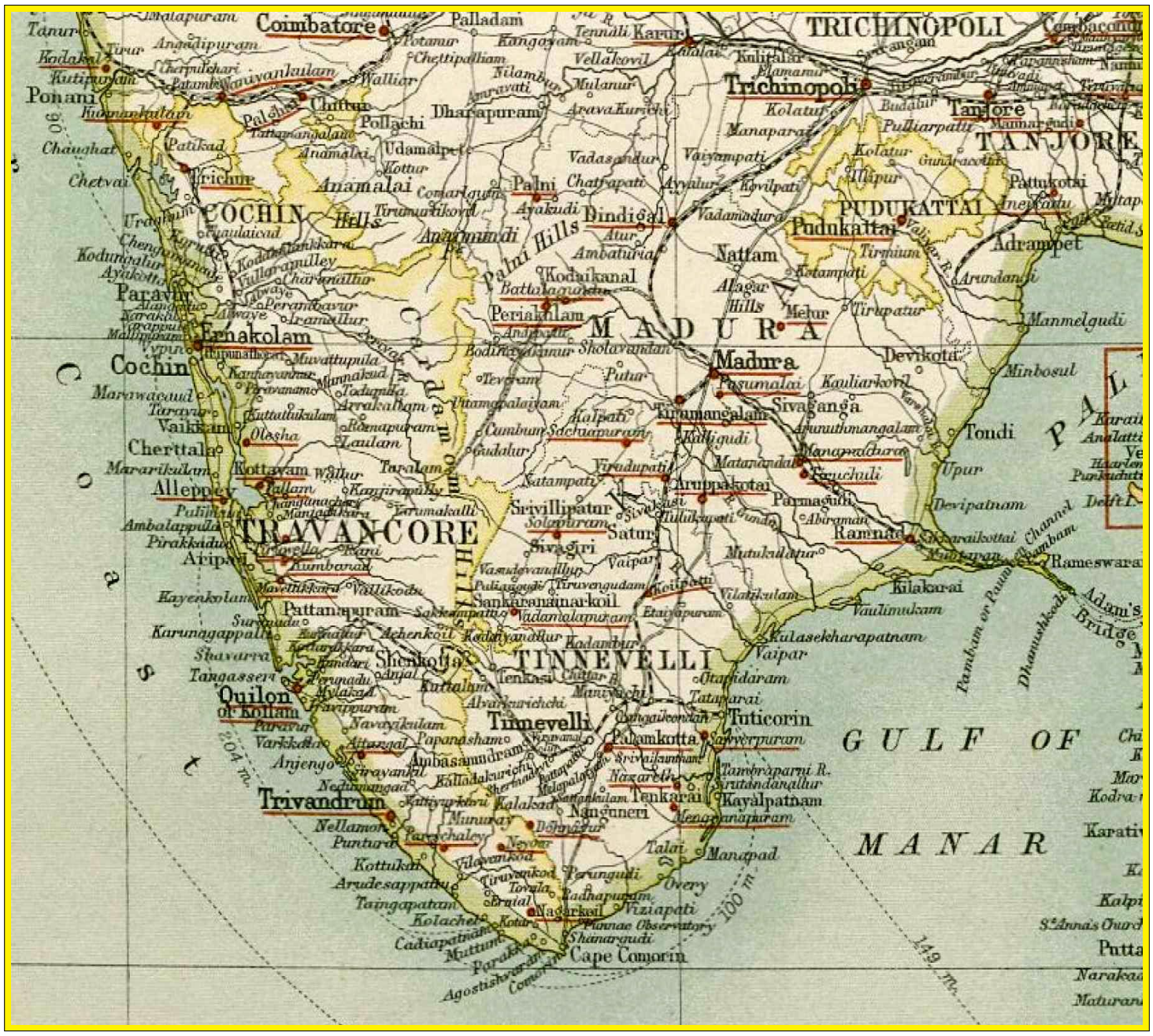
Wantchekon, L., N. Novta, and M. Klasnja (2013). Education and Human Capital Externalities: Evidence from Colonial Benin. Unpublished manuscript. [5]

Wilson, R. S. (1928). *The Indirect Effects of Christian Missions in India*. London, U.K.: James Clarke & Co. [26]

Wood, A. S. (1969). *Captive to the Word. Martin Luther: Doctor of Sacred Scripture*. Exeter, U.K.: Paternoster Press. [2]

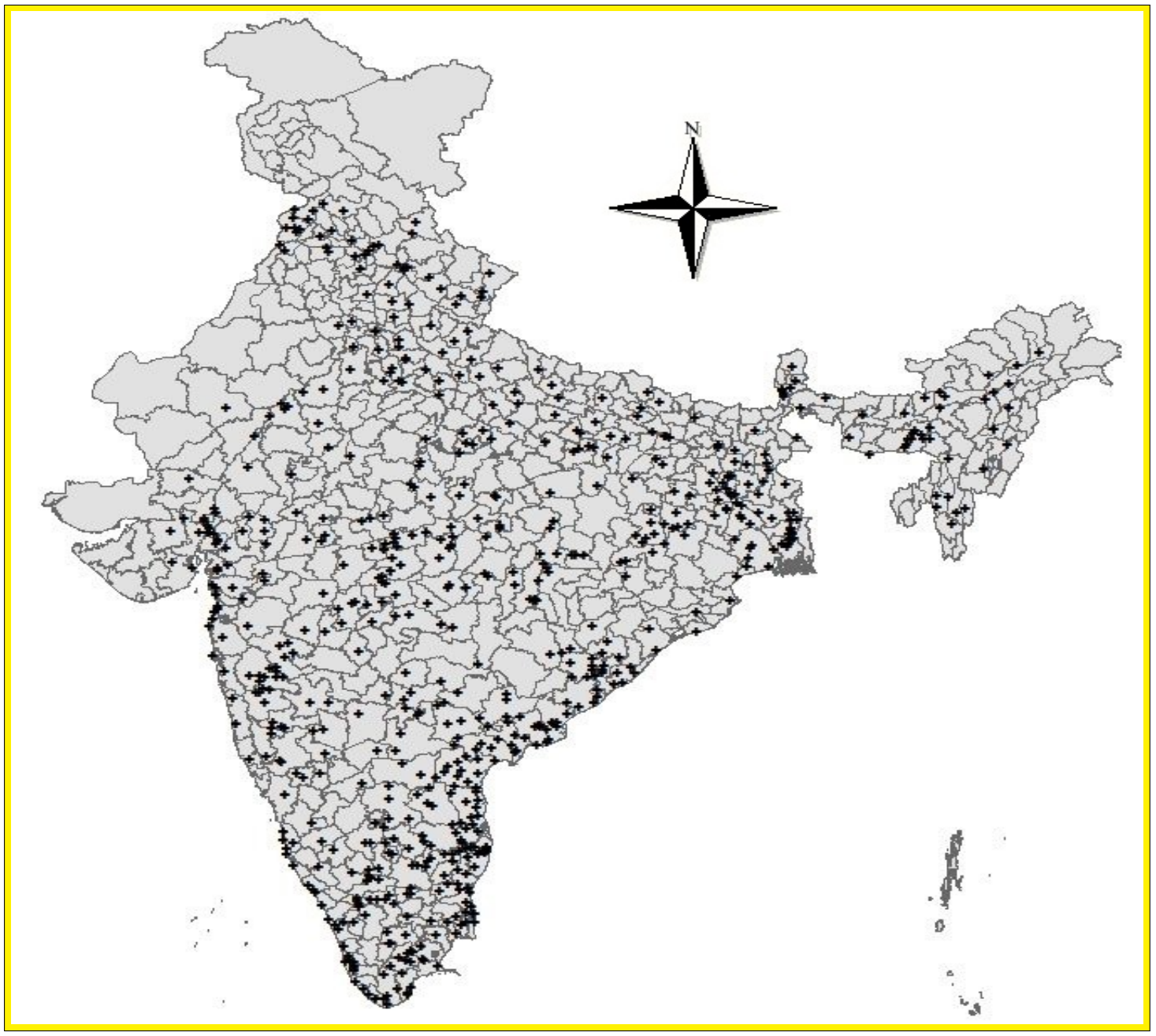
Figures

Figure 1: Map's detail from the "Statistical Atlas of Christian Missions" (1910)



Note: towns with Protestant missions as of 1908 are underlined in red.

Figure 2: Historical Protestant Missions in India

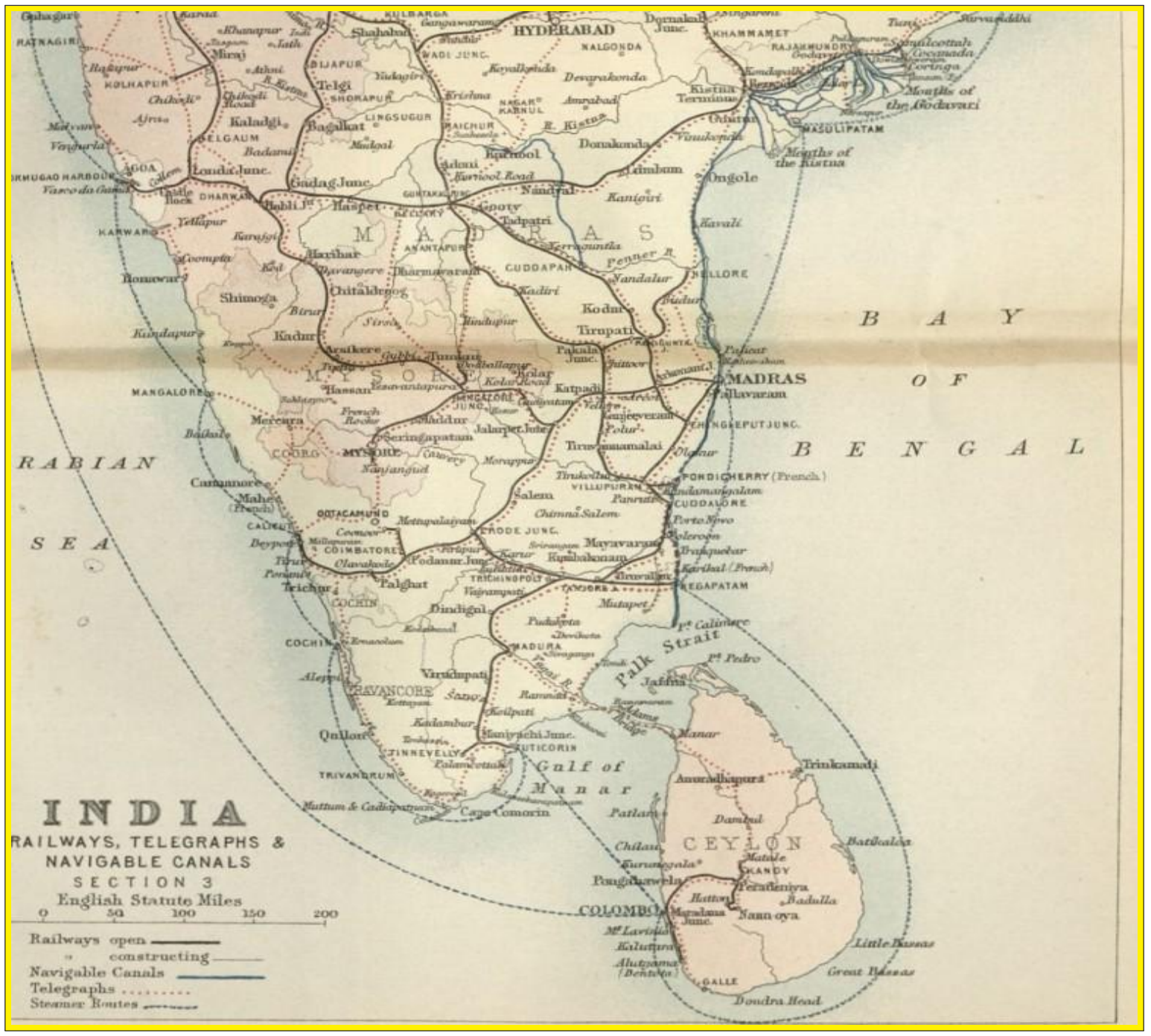


Note: this figure combines information about the location of Protestant missions in India as of 1908 contained in the "Statistical Atlas of Christian Missions" (little crosses) with a map showing the administrative boundaries of the Indian districts in 2001.

Figure 3: Political division of India in 1837

Note: this map is from the "Imperial Gazetteer of India" (1909, vol. XXVI) and shows the political division of India as of 1837. Areas in pink were under direct British control; regions in yellow and green were ruled by, respectively, Hindu and Muslim princes.

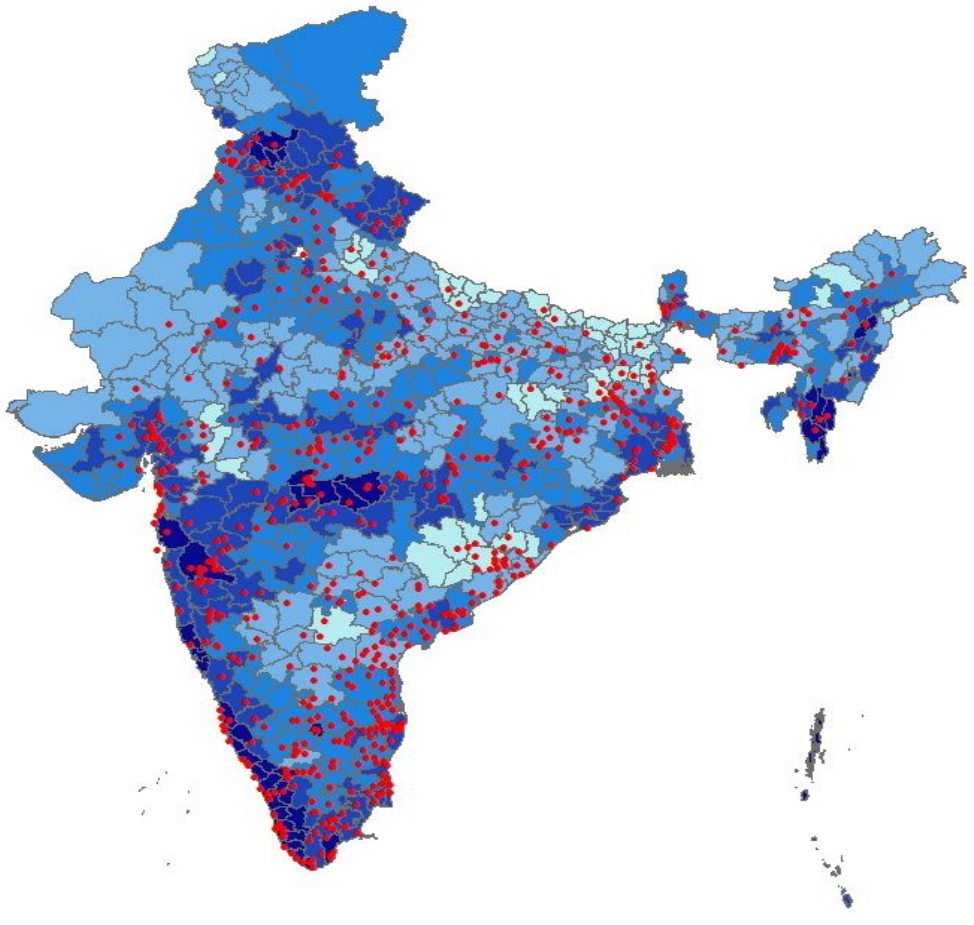
Figure 4: Railways in India as of 1891



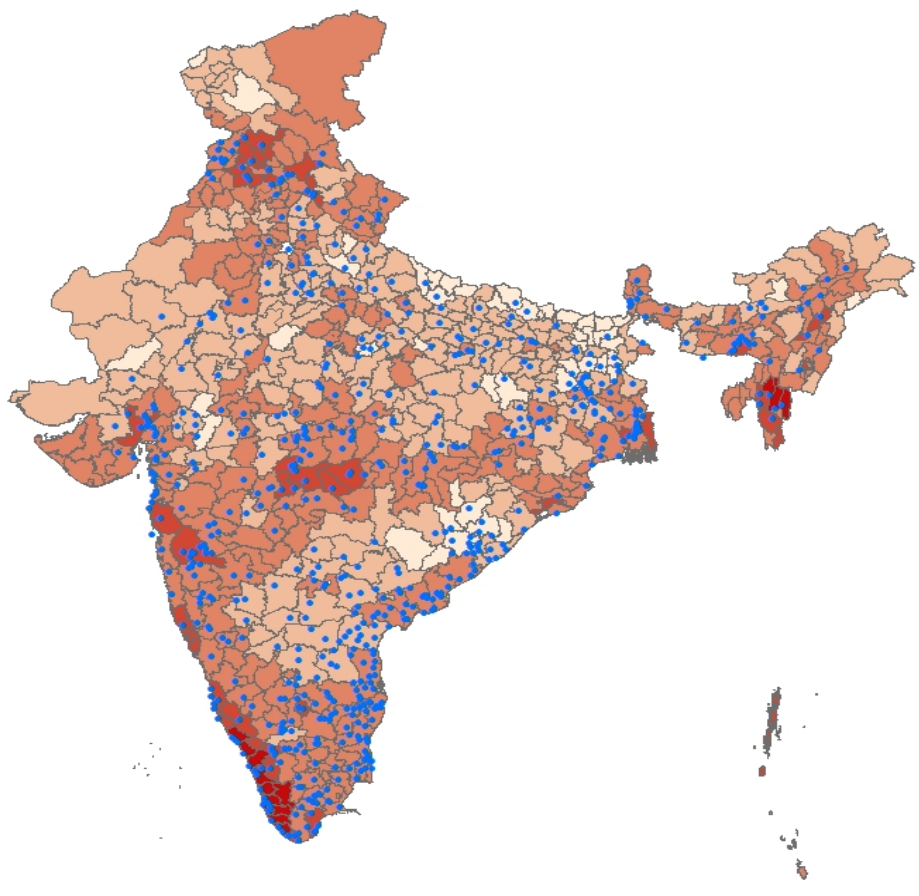
Note: section of Plate 20 from the "Constable's Hand Atlas of India" (1893) showing the railways network in India as of 1891.

Figure 5: Missions and Literacy Rates

(A) Population Literacy Rate



(B) Female Literacy Rate



(C) Gender Parity Index

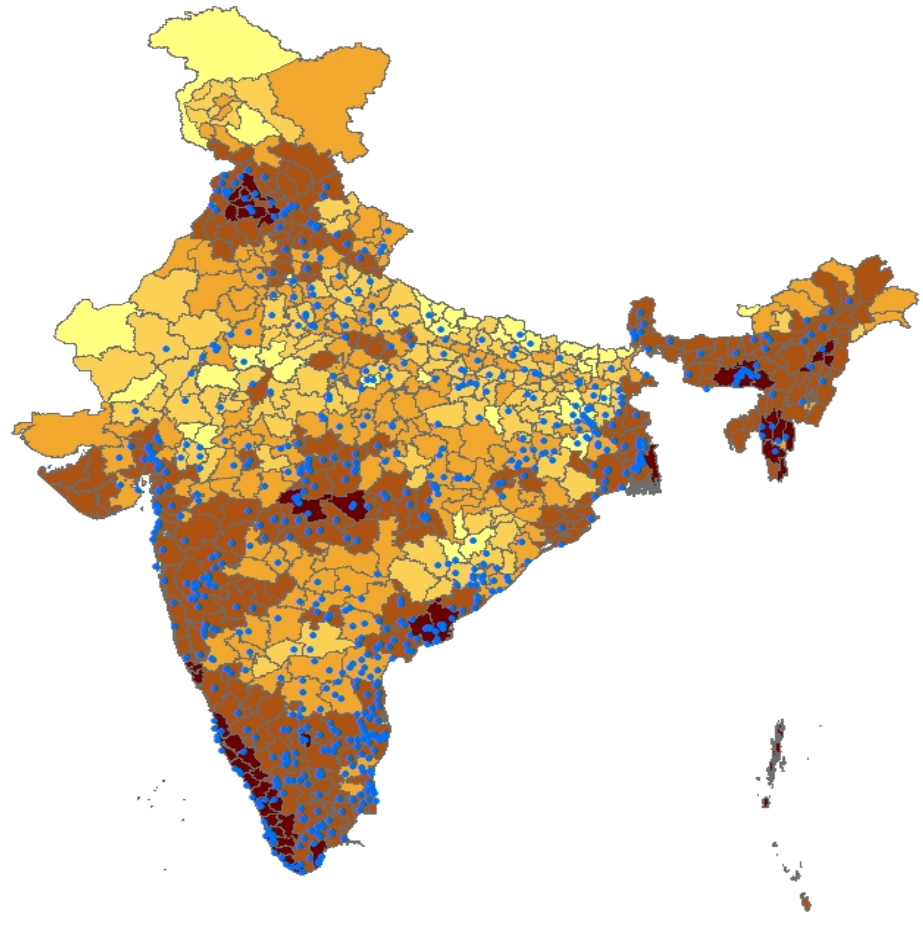
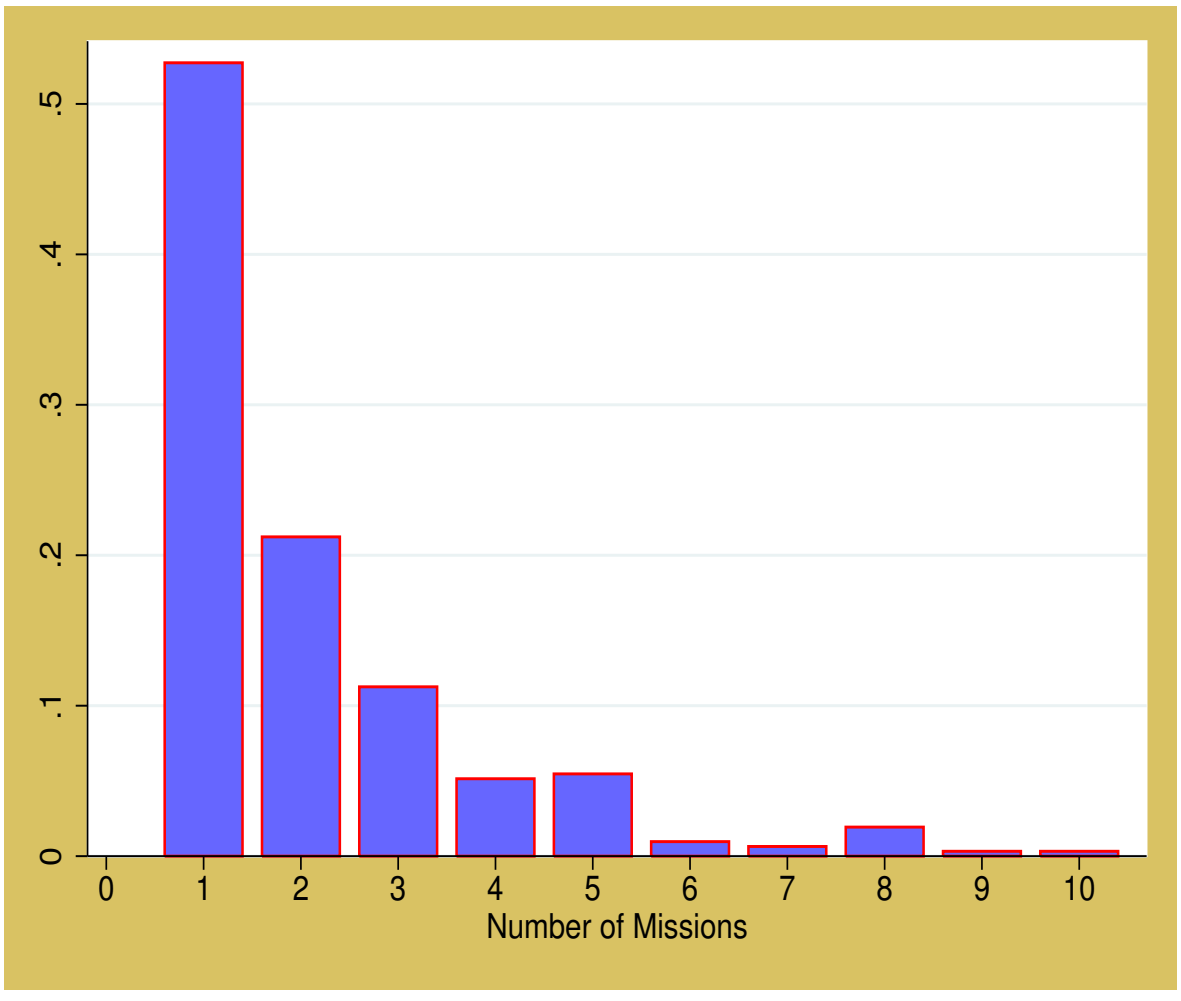


Figure 6: Map's detail from the "Atlas Hierarchicus" (1913).

Note: towns with Catholic missions as of 1911 are underlined.

Figure 7: Districts' Number of Protestant Missions



Note: conditional on the number of Protestant missions being positive.

Tables

Table 1: Variables' Description

Outcome Variables:	
Total Population Literacy Rate	Total population literacy rate. Source: Census (2001).
Female Literacy Rate	Female population literacy rate. Source: Census (2001).
GPI	Gender parity index defined as $\frac{\text{Female Lit. Rate}}{\text{Male Lit. Rate}}$. Source: author's calculation based on data from Census (2001).
Current Controls:	
Urban	Share of population living in urban areas. Source: Census (2001).
Electricity	Share of district's villages that have access to electric power. Source: Census (2001).
Medical	Share of district's villages with medical facilities. Source: Census (2001).
Bus Stop	Share of district's villages with a bus stop. Source: Census (2001).
Paved Road	Share of district's villages with a paved entry road. Source: Census (2001).
Primary Schools	Share of district's villages with a primary school. Source: Census (2001).
Secondary Schools	Share of district's villages with a secondary school. Source: Census (2001).
College	Share of district's villages with a college/university. Source: Census (2001).
% Christians	Population share of Christians. Source: Census (2001).
Geographic Controls:	
Latitude	Latitude (in radians) of the centroid of the district. Source: author's calculation.
Longitude	Longitude (in radians) of the centroid of the district. Source: author's calculation.
Latid. x Longit.	Latitude x Longitude of the centroid of the district. Source: author's calculation.
Coast	Dummy = 1 if coastal district. Source: author's calculation.
Rivers	Total kilometers of rivers per squared kilometer in the district. Source: author's calculation based on data from the "Digital Chart of the World" (http://www.diva-gis.org/gdata).
City Distance	Minimum distance (in kilometers) of district's centroid from one of the ten major cities in India (Mumbai, Delhi, Bangalore, Hyderabad, Ahmedabad, Chennai, Kolkata, Surat, Pune, Jaipur). Source: author's calculation.
Altitude	Average altitude (in meters) of the district. Source: author's calculation based on data from the "CGIAR Consortium for Spatial Information" (http://srtm.csi.cgiar.org/)
Historical Controls:	
1901 Density	Population per squared mile. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 Urban	Share of population living in urban areas. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 No-Agr. Pop.	Share of labor force not employed in agricultural sector. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 Infirmities	Share of population affected by infirmity (insanity, deafness-mutism, blindness, leprosy). Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 Muslims	Population share of Muslims. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 Total Literacy Rate	Total population literacy rate. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 Female Literacy Rate	Female population literacy rate. Source: Census (1901), "Imperial Gazetteer of India" (1909) and various District Gazetteers.
1901 GPI	Gender parity index defined as $\frac{\text{Female Lit. Rate}}{\text{Male Lit. Rate}}$. Source: Census (1901).
1837 British	Dummy = 1 if district belongs to a territory under direct British rule in 1837. Source: "Imperial Gazetteer of India" (1909), vol. XXVI.
1837 Hindu	Dummy = 1 if district belongs to a territory ruled by Hindu prince in 1837. Source: "Imperial Gazetteer of India" (1909), vol. XXVI.
1837 Muslim	Dummy = 1 if district belongs to a territory ruled by Muslim prince in 1837. Source: "Imperial Gazetteer of India" (1909), vol. XXVI.
1891 Railroads	Total kilometers of railroads per squared kilometer in the district. Source: author's calculation based on maps from the "Constable's Hand Atlas of India" (1893), plates
Missions:	
Protestant Mission	Dummy = 1 if Protestant missions in the district as of 1908. Source: author's calculation based on information contained in the "Statistical Atlas of Christian Missions" (1910).
Protestant Density	Number of Protestant missions (as of 1908) per squared mile: $\frac{\#\text{missions in district } i}{\text{area (in squared miles) of district } i}$. Source: author's calculation based on information contained in the "Statistical Atlas of Christian Missions" (1910).
Protestant Distance	District's average minimum distance (in km) from historical Protestant mission. Source: author's calculation based on information contained in the "Statistical Atlas of Christian Missions" (1910). See text for more details.

Table 2: Descriptive Statistics

	Mean	S.D.	Median	25th P.tile	75th P.tile	Min.	Max.	Obs.
Outcome Variables:								
Total Population Literacy Rate	0.63	0.13	0.63	0.55	0.72	0.30	0.94	515
Female Literacy Rate	0.51	0.12	0.51	0.41	0.62	0.20	0.93	515
GPI	0.68	0.12	0.69	0.59	0.76	0.39	1.07	515
Current Controls:								
% Christians	0.06	0.18	0.003	0.008	0.017	0	0.96	515
Urban	0.21	0.15	0.17	0.10	0.27	0	0.94	515
% of district's villages with:								
Electricity	0.80	0.24	0.92	0.68	0.99	0.05	1	515
Medical Facilities	0.39	0.25	0.33	0.19	0.54	0.022	1	515
Bus Stop	0.44	0.31	0.34	0.16	0.75	0.017	1	515
Paved Road	0.60	0.25	0.58	0.38	0.83	0.10	1	515
Primary Schools	0.83	0.15	0.88	0.77	0.94	0.22	1	515
Secondary Schools	0.20	0.19	0.14	0.08	0.26	0.002	1	515
College	0.01	0.02	0.005	0.002	0.01	0	0.16	515
Geographic Controls:								
Latitude (radians)	0.41	0.09	0.43	0.36	0.47	0.14	0.60	515
Longitude (radians)	1.41	0.11	1.39	1.33	1.48	1.22	1.68	515
Latit. x Longit.	0.58	0.15	0.62	0.50	0.69	0.19	0.83	515
Coast	0.10	0.30	0	0	1	0	1	515
Rivers (km/km^2)	0.10	0.06	0.09	0.05	0.13	0	0.32	515
Altitude (meters)	473	708	246	118	482	5	5,062	515
City Distance (km)	357	197	344	202	502	22	1,022	515
Historical Controls:								
1901 Density (Pop/sqm)	275	216	201	107	418	5	985	515
1901 Urban	0.09	0.07	0.08	0.03	0.12	0	0.51	516
1901 Population Literacy Rate	0.045	0.081	0.035	0.025	0.052	0.006	0.24	515
1901 Female Literacy Rate	0.005	0.008	0.002	0.001	0.005	0.004	0.09	515
1901 GPI	0.051	0.046	0.037	0.026	0.057	0.009	0.32	515
1901 % No-Agr. Pop.	0.17	0.08	0.17	0.11	0.23	0.015	0.53	515
1901 Infirmities	0.004	0.006	0.002	0.002	0.004	0	0.065	515
1901 Muslims	0.16	0.21	0.09	0.04	0.22	0	0.88	515
1837 British	0.58	0.49	1	0	1	0	1	515
1837 Hindu	0.33	0.46	0	0	1	0	1	515
1837 Muslim	0.09	0.27	0	0	1	0	1	515
1891 Railways	0.0017	0.0039	0.0009	0	0.020	0	0.0283	515
Missions								
Protestant Mission	0.55	0.49	1	0	1	0	1	515
Protestant Density (x100)	0.07	0.09	0.03	0	0.10	0	0.60	515
Protestant Distance (km)	45.7	38.8	34.3	24.9	51.6	10.8	304	515

Table 3: Historical Protestant Missions and Current Literacy Outcomes

	[1]	[2]	[3]	[4]
Dep. Variable: Total Population Literacy Rate				
Protestant Mission = 1	0.0553***	0.0333***	0.0306***	0.0253***
	(0.0166)	(0.0088)	(0.0083)	(0.0079)
Dep. Variable: Female Literacy Rate				
Protestant Mission = 1	0.0718***	0.0413***	0.0373***	0.0301***
	(0.0194)	(0.0106)	(0.0091)	(0.0083)
Dep. Variable: Gender Parity Index				
Protestant Mission = 1	0.0567***	0.0301***	0.0272***	0.0210***
	(0.0139)	(0.0081)	(0.0069)	(0.0065)
Observations	515	515	515	515
State FE	NO	YES	YES	YES
Geographic Controls	NO	NO	YES	YES
Historical Controls	NO	NO	NO	YES

Table 3, organized in three panels, reports OLS estimates associating modern literacy outcomes (respectively: total population literacy rate, female literacy rate and gender parity index) to the presence of Protestant missions as of 1908. The unit of analysis is a district.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 4: Altonji Ratios

Dependent Variable ↓	Full (α^F)	Restricted (α^R)	Altonji-Elder-Taber (AET) Ratio
Population Literacy Rate	0.0253	0.0333	3.16
Female Literacy Rate	0.0301	0.0413	2.68
Gender Parity Index	0.0210	0.0301	2.31

AET Ratio = $|\alpha^F/(\alpha^R - \alpha^F)|$. The "restricted" and "full" regressions correspond, respectively , to the specification of column [2] and column [4] in Table 3.

Table 5: Geographical and Historical Characteristics within Groups of Neighbor Districts

	Dependent Variable is:												
	Distance Major City	Coast Dummy	Rivers	Mean Alt.	Urban 1901	Pop.Density 1901	No Agric. 1901	Muslims 1901	Infirmities 1901	Railroads 1891	Tot. Liter. 1901	Female Liter. 1901	GPI 1901
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
Protestant Mission = 1	-0.425	0.0269	-0.0137	-33.41	0.0038	0.0004	-0.0037	-0.00623	-0.0010	0.0003	-0.003	-0.0008	0.0029
	(5.667)	(0.0369)	(0.00824)	(43.68)	(0.00672)	(0.0225)	(0.00631)	(0.0119)	(0.00167)	(0.000575)	(0.006)	(0.00170)	(0.00329)
Neighbor Districts FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	250	250	250	250	250	250	250	250	250	250	250	250	250

Table 5 reports OLS estimates associating several geographic and historical characteristics to the presence of Protestant missions as of 1908. The unit of analysis is a district. The sample consists of neighbor districts differing in the sense that one district contains historical Protestant missions while the adjacent one does not. The estimation includes (within State) neighbor district fixed effects. Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 6: Historical Protestant Missions and Current Literacy Outcomes within Neighbor Districts in the Same State

Dep. Var. →	Total Pop. Literacy Rate [1]	Female Literacy Rate [2]	Gender Parity Index [3]
Protestant Mission = 1	0.0248**	0.0293**	0.0152*
	(0.0102)	(0.0112)	(0.0084)
Neighbor Districts FE	YES	YES	YES
Geographic Controls	YES	YES	YES
Historical Controls	YES	YES	YES
Observations	250	250	250

Table 6 reports OLS estimates associating modern literacy outcomes (total population literacy rate, female literacy rate and gender parity index) to the presence of Protestant missions as of 1908. The unit of analysis is a district. The estimation sample consists of neighbor districts differing in the sense that one district contains historical Protestant missions while the adjacent one does not.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.

Table 7: Using Catholic Missions as Control Group

Dep. Var. →	Tot. Lit. Rate [1]	Female Lit. Rate [2]	GPI [3]
Protestant Mission = 1	0.0263***	0.0306***	0.0196***
	(0.008)	(0.009)	(0.006)
Catholic Mission = 1	-0.006	-0.003	0.009
	(0.006)	(0.007)	(0.006)
State FE	YES	YES	YES
Geographic Controls	YES	YES	YES
Historical Controls	YES	YES	YES
Observations	515	515	515

Table 7, organized in three panels, reports OLS estimates associating modern literacy outcomes (respectively: total population literacy rate, female literacy rate and gender parity index) to the presence of Protestant missions as of 1908 (measured by a binary indicator equal to one if there were Protestant missions in the district's territory as of 1908 and zero otherwise). The unit of analysis is a district.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 8: Historical Protestant Missions and Literacy of Scheduled Caste Population

	[1]	[2]	[3]	[4]	[5]	[6]
Protestant Mission = 1	0.0474**	0.0335**	0.0322***	0.0254**	0.0265**	0.0198*
	(0.0224)	(0.0121)	(0.0115)	(0.0109)	(0.0117)	(0.0101)
Observations	471	471	471	471	471	471
State FE	NO	YES	YES	YES	YES	YES
Geographic Controls	NO	NO	YES	YES	YES	YES
Historical Controls	NO	NO	NO	YES	YES	YES
Catholic Mission	NO	NO	NO	NO	YES	NO
School Supply	NO	NO	NO	NO	NO	YES
Development Controls	NO	NO	NO	NO	NO	YES
% Christians	NO	NO	NO	NO	NO	YES

Table 8 reports the OLS estimates associating the current literacy rate of the district's Scheduled Caste population to the presence of Protestant missions as of 1908 (measured by a binary indicator equal to one if there were Protestant missions in the district's territory as of 1908 and zero otherwise). The unit of analysis is a district. The analysis excludes the districts in the states of Nagaland and Andaman and Nicobar Islands as they lack data about the scheduled caste population in 2001. Moreover I include in the sample only those districts where individuals belonging to scheduled castes represents at least one percent of the district's population.
The specification in column [5] uses Catholic missions as control group by adding a binary indicator equal to one if the district was populated by Catholic missions and zero otherwise.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
"Development Controls" include: degree of urbanization, share of district's villages with access to electricity, medical facilities, bus stop, paved entry road.
"School Supply" includes three variables measuring the district's share of villages with a primary school, a secondary school, a college. "% Christians" measures the district's share of Christian population.
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 9: Historical Protestant Missions and Current Literacy Outcomes: Channels of Transmission

Dep. Var. →	Total Pop. Literacy Rate	Female Literacy Rate	Gender Parity Index
	[1]	[2]	[3]
Protestant Mission = 1	0.0159**	0.0200**	0.0140*
$[\alpha^{\text{FULL}}]$	(0.0067)	(0.0075)	(0.0069)
% of district's villages with:			
Primary School	-0.0247	-0.0268	-0.0244
	(0.0368)	(0.0458)	(0.0406)
Secondary School	0.125**	0.112*	0.0255
	(0.0613)	(0.0600)	(0.0346)
College	0.746*	1.002**	0.917***
	(0.322)	(0.378)	(0.244)
% Urban	0.196***	0.235***	0.165***
	(0.0276)	(0.0315)	(0.0257)
% of district's villages with:			
Electricity	0.149***	0.145***	0.0884**
	(0.0416)	(0.0468)	(0.0405)
Medical Facilities	-0.0469	-0.0606	-0.0400
	(0.0417)	(0.0452)	(0.0301)
Bus Stop	-0.0089	0.0241	0.0406
	(0.0506)	(0.0590)	(0.0488)
Paved Entry Road	-0.0083	-0.0165	-0.0138
	(0.0347)	(0.0362)	(0.0243)
% Christians	0.0345	0.0721	0.0933
	(0.0440)	(0.0502)	(0.0575)
State FE	YES	YES	YES
Geographic Controls	YES	YES	YES
Historical Controls	YES	YES	YES
Observations	515	515	515
α^{BASELINE}	0.0253***	0.0301***	0.0210***
Contribution to $\Delta\alpha$ by:			
Schools Supply	-0.0006	-0.0006	0.0001
Current Development	0.0093**	0.0099**	0.0058*
Christian Share	0.0005	0.0008	0.001

Table 9, organized in three panels, reports state fixed effects OLS estimates associating modern literacy outcomes (respectively: total population literacy rate, female literacy rate and gender parity index) to the presence of Protestant missions as of 1908 (measured by a binary indicator equal to one if there were Protestant missions in the district's territory as of 1908 and zero otherwise). The unit of analysis is a district.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.
The bottom part of the table reports the degree to which three different groups of variables reduce (or increase) the size of the effect of Protestant missions on literacy outcomes. α^{BASELINE} refers to the estimated coefficients in Table 3, $\Delta\alpha = \alpha^{\text{BASELINE}} - \alpha^{\text{FULL}}$.
"School Supply": primary school, secondary school, college. "Current Development": urban, electricity, medical facilities, bus stop, paved entry road. "Christian Share": % Christians. The decomposition is based on a method developed by Gelbach (2009). See Appendix B for more details.

A Additional Tables

Table A.1: Differences Between "Protestant" and "non-Protestant" Neighbor Districts

Variable ↓	Protestant Mission = 0	Protestant Mission = 1	Difference
City Distance	391.77	391.05	0.72
Coast	0.01	0.04	-0.03*
Rivers	0.11	0.09	0.015*
Altitude	495.4	441.9	53.4
1901 Urban	0.077	0.080	-0.003
1901 Density	0.251	0.263	-0.011
1901 % No-Agr. Pop.	0.169	0.170	-0.0004
1901 Muslims	0.197	0.179	0.018
1901 Infirmities	0.0041	0.0040	0.0001
1901 Railways	0.0015	0.0019	-0.0004
1901 Population Literacy Rate	0.053	0.041	0.011
1901 Female Literacy Rate	0.0043	0.0045	-0.0001
1901 GPI	0.042	0.050	-0.008*

Table A.1 reports averages of the listed variables differentiating between "Protestant" and "non-Protestant" districts belonging to the sample of neighbor districts. ***, ** and * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table A.2: Robustness Check: Excluding Outliers and Districts under the Portuguese Control

	[1]	[2]	[3]	[4]	Obs.
DV: Total Population Literacy Rate					
Excluding:					
Outliers 1%	0.0475***	0.0285***	0.0243***	0.0185**	506
	(0.0148)	(0.007)	(0.008)	(0.007)	
Outliers 2%	0.0429***	0.0282***	0.0241***	0.0189**	496
	(0.0154)	(0.008)	(0.008)	(0.008)	
Outliers 5%	0.0412***	0.0325***	0.0289***	0.0218***	466
	(0.014)	(0.011)	(0.008)	(0.008)	
Portuguese Districts	0.0499***	0.0330***	0.0310***	0.0250***	502
	(0.0160)	(0.009)	(0.008)	(0.008)	
DV: Female Literacy Rate					
Excluding:					
Outliers 1%	0.0657***	0.0401***	0.0346***	0.0279***	506
	(0.0175)	(0.0104)	(0.008)	(0.008)	
Outliers 2%	0.0575***	0.0363***	0.0308***	0.0245***	496
	(0.0175)	(0.0101)	(0.009)	(0.008)	
Outliers 5%	0.0556***	0.0384***	0.0346***	0.0260***	466
	(0.0173)	(0.0114)	(0.009)	(0.008)	
Portuguese Districts	0.0648***	0.0409***	0.0380***	0.0307***	502
	(0.0185)	(0.0107)	(0.0090)	(0.0081)	
DV: Gender Parity Index					
Excluding:					
Outliers 1%	0.0523***	0.0293***	0.0251***	0.0204***	506
	(0.0121)	(0.008)	(0.007)	(0.006)	
Outliers 2%	0.0482***	0.0290***	0.0246***	0.0200***	496
	(0.0119)	(0.008)	(0.007)	(0.006)	
Outliers 5%	0.0386***	0.0250**	0.0210**	0.0168**	466
	(0.0130)	(0.009)	(0.008)	(0.006)	
Portuguese Districts	0.0523***	0.0302***	0.0281***	0.0226***	502
	(0.0135)	(0.008)	(0.007)	(0.006)	
DV: SC Literacy Rate					
Excluding:					
Outliers 1%	0.0465**	0.0349***	0.0342***	0.0280**	466
	(0.021)	(0.011)	(0.011)	(0.010)	
Outliers 2%	0.0373*	0.0296**	0.0285**	0.0212**	460
	(0.019)	(0.012)	(0.012)	(0.010)	
Outliers 5%	0.0273	0.0286**	0.0258*	0.0192*	438
	(0.019)	(0.013)	(0.013)	(0.010)	
Portuguese Districts	0.0418*	0.0331**	0.0322**	0.0246**	458
	(0.022)	(0.012)	(0.012)	(0.012)	
State FE	NO	YES	YES	YES	
Geographic Controls	NO	NO	YES	YES	
Historical Controls	NO	NO	NO	YES	

Table A.2, organized in three panels, reports OLS estimates associating modern literacy outcomes (respectively: total population literacy rate, female literacy rate and gender parity index) to the presence of Protestant missions as of 1908 (measured by a binary indicator equal to one if there were Protestant missions in the district's territory as of 1908 and zero otherwise). The unit of analysis is a district.
In each panel, I perform the estimation of Equation 1 dropping observations falling in the extremes of the distribution of the dependent variables and excluding those districts that belong to territories fallen under the Portuguese control during the colonial period.
Column [2] adds state specific fixed effects. Column [3] includes "Geographic Controls": latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer. Column [4] includes "Historical Controls": share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table A.3: Using Historical Missions Density as Measure of Missionary Influence

	Full Sample				Neighbor Districts Sample
	[1]	[2]	[3]	[4]	[5]
Dep. Variable: Total Population Literacy Rate					
Protestant Density (x100)	0.427***	0.309***	0.258***	0.180***	0.224**
	(0.079)	(0.042)	(0.049)	(0.048)	(0.098)
Change due to a 1 s.d. increase of historical Protestant Density	[3.4 p.p.]	[2.5 p.p.]	[2.1 p.p.]	[1.4 p.p.]	[1.8 p.p.]
Dep. Variable: Female Literacy Rate					
Protestant Density (x100)	0.554***	0.365***	0.299***	0.194***	0.231**
	(0.099)	(0.049)	(0.055)	(0.052)	(0.102)
Change due to a 1 s.d increase of historical Protestant Density	[4.4 p.p.]	[3.0 p.p.]	[2.4 p.p.]	[1.6 p.p.]	[1.8 p.p.]
Dep. Variable: Gender Parity Index					
Protestant Density (x100)	0.432***	0.244***	0.193***	0.103**	0.130
	(0.078)	(0.041)	(0.041)	(0.043)	(0.077)
Change due to a 1 s.d. increase of historical Protestant Density	[3.4 p.]	[1.9 p.]	[1.5 p.]	[0.8 p.]	[1.0 p.]
Observations	515	515	515	515	250
Dep. Variable: SC Literacy Rate					
Protestant Density (x100)	0.367***	0.338***	0.292***	0.235***	0.223***
	(0.105)	(0.049)	(0.048)	(0.052)	(0.095)
Change due to a 1 s.d. increase of historical Protestant Density	[2.9 p.p.]	[2.7 p.p.]	[2.3 p.p.]	[1.9 p.p.]	[1.8 p.p.]
Observations	471	471	471	471	232
State FE	NO	YES	YES	YES	YES
Neighbor Districts FE	-	-	-	-	YES
Geographic Controls	NO	NO	YES	YES	YES
Historical Controls	NO	NO	NO	YES	YES

Table A.3, organized in three panels, reports OLS estimates associating modern literacy outcomes (respectively: total population literacy rate, female literacy rate and gender parity index) to the density of Protestant missions as of 1908. The unit of analysis is a district.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.
In squared brackets is reported the change in the outcome variable due to a one standard deviation increase of historical Protestant density (p.p. = percentage points; p. = points).

Table A.4: Introducing a Buffer around the Districts to Compute the Historical Missions Density

	Length of buffer's radius		
Dependent Variable ↓	10 km [1]	20 km [2]	40 km [3]
Population Literacy Rate	0.137***	0.111***	0.051***
	(0.050)	(0.036)	(0.017)
Female Literacy Rate	0.155***	0.121***	0.062***
	(0.053)	(0.038)	(0.019)
Gender Parity Index	0.090**	0.058**	0.023
	(0.039)	(0.028)	(0.016)
Observations	515	515	515
SC Literacy Rate	0.202***	0.149***	0.068**
	(0.059)	(0.041)	(0.027)
Observations	471	471	471
State FE	YES	YES	YES
Geographic Controls	YES	YES	YES
Historical Controls	YES	YES	YES

Table A.4 reports the estimated coefficients for the "Protestant Density" variable when this is computed by drawing around each district a buffer with a radius of a varying length (10, 20 or 40 km). The unit of analysis is a district.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer, mean altitude of the district.
"Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table A.5: Using District's Average Minimum Distance from Historical Missions as Measure of Missionary Influence

	Full Sample				Protestant *OR* Catholic				Protestant *AND* Catholic			
	Tot. [1]	Fem. [2]	GPI [3]	SC [4]	Tot. [5]	Fem. [6]	GPI [7]	SC [8]	Tot. [9]	Fem. [10	GPI [11]	SC [12]
Min. Distance from Protestant Mission (/100)	-0.096***	-0.093***	-0.036**	-0.127***	-0.174***	-0.178***	-0.094**	-0.252***	-0.236**	-0.301**	-0.194**	-0.291***
	(0.020)	(0.021)	(0.015)	(0.025)	(0.040)	(0.044)	(0.034)	(0.047)	(0.106)	(0.117)	(0.091)	(0.083)
Min. Distance from Catholic Mission (/100)	0.022	0.021	0.004	0.025	0.002	-0.005	-0.015	-0.002	-0.004	-0.132	-0.075	-0.136
	(0.013)	(0.016)	(0.013)	(0.019)	(0.018)	(0.022)	(0.024)	(0.024)	(0.072)	(0.074)	(0.052)	(0.097)
Observations	515	515	515	471	311	311	311	295	118	118	118	117
State FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Geographic Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Historical Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table A.5 reports OLS estimates when the degree of historical Protestant influence is measured as the district's average minimum distance (in kilometers) from a Protestant mission (see text for details about the construction of this variable). The unit of analysis is a district.
"Tot." = total population literacy rate; "Fem." = female literacy rate; "GPI" = gender parity index; "SC" = literacy rate of Scheduled Caste population.
"Geographic Controls" include: latitude of the centroid of the district, longitude of the centroid of the district, interaction of latitude and longitude, distance of the centroid of the district to a major Indian city, a dummy equal to one if coastal district, the total length (in kilometers) of rivers per squared kilometer. "Historical Controls" include: share of population living in urban areas, population density, literacy rate, share of non-agricultural labor force, share of infirm population, share of Muslim population (all these variables measured in 1901), railroad network in 1891 and two dummy variables indicating if, in 1837, the district was ruled by a Muslim or by a Hindu prince (excluded category: district was under direct British control).
"Protestant *OR* Catholic" refers to the sample of districts that were inhabited by either Protestant or Catholic missions. "Protestant *AND* Catholic" refers to the sample of districts that had both Protestant and Catholic missions.
Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, ** and * mean statistical significance at, respectively, 1, 5 and 10 percent level.

B Decomposition Method (Gelbach [2009])

Assume that the baseline model has the form

$$Y = \alpha_0^B + \alpha_T^B T + \alpha_1^B X_1 + \epsilon^B \tag{5}$$

where Y is an outcome variable, T is the regressor of interest, X_1 is a group of covariates and ϵ is an error term.

Assume that we then estimate a "full" regression where we include additional groups of regressors X_2, X_3 and X_4 (in my case X_2 = school supply controls; X_3 = development controls; X_4 = Christian share). That is the full model is:

$$Y = \alpha_0^F + \alpha_T^F T + \alpha_1^F X_1 + \alpha_2^F X_2 + \alpha_3^F X_3 + \alpha_4^F X_4 + \epsilon^F \tag{6}$$

Gelbach (2009) develops a method, based on the Oaxaca-Blinder decomposition, that allows to quantify the contribution to $\Delta\widehat{\alpha} = (\widehat{\alpha}_T^B - \widehat{\alpha}_T^F)$ of each group of covariates X_2, X_3 and X_4. The relationship between $\widehat{\alpha}_T^B$ and $\widehat{\alpha}_T^F$ is expressed as:

$$\widehat{\alpha}_T^B = \widehat{\alpha}_T^F + \left[\sum_{j=1}^{N_2} \widehat{\theta}_2 \widehat{\alpha}_2^F\right] + \left[\sum_{j=1}^{N_3} \widehat{\theta}_3 \widehat{\alpha}_3^F\right] + \left[\sum_{j=1}^{N_4} \widehat{\theta}_4 \widehat{\alpha}_4^F\right] \tag{7}$$

where $\widehat{\alpha}_k^F$ $(k = 1, ..., 4)$ are defined in Equation 6; N_k is the number of covariates in each of the k-groups; $\widehat{\theta}_k$ are defined by the auxiliary regression

$$T = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \eta \tag{8}$$

Rearranging Equation 7 we obtain

$$\left(\widehat{\alpha}_T^B - \widehat{\alpha}_T^F\right) = \left[\sum_{j=1}^{N_2} \widehat{\theta}_2 \widehat{\alpha}_2^F\right] + \left[\sum_{j=1}^{N_3} \widehat{\theta}_3 \widehat{\alpha}_3^F\right] + \left[\sum_{j=1}^{N_4} \widehat{\theta}_4 \widehat{\alpha}_4^F\right] \tag{9}$$

where each term in squared brackets represents the contribution of each group of covariates in explaining $\Delta\widehat{\alpha}$.