Essays in Development Economics

Author: Federico Mantovanelli

Persistent link: http://hdl.handle.net/2345/3793

This work is posted on eScholarship@BC,
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2014

Copyright is held by the author, with all rights reserved, unless otherwise noted.

Boston College

The Graduate School of Arts and Sciences

Department of Economics

ESSAYS IN DEVELOPMENT ECONOMICS

a dissertation

by

FEDERICO G. MANTOVANELLI

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May, 2014

© copyright by FEDERICO G. MANTOVANELLI
2014

ESSAYS IN DEVELOPMENT ECONOMICS

ABSTRACT

by

FEDERICO G. MANTOVANELLI

Dissertation Committee:

SCOTT FULFORD (Co-chair)

MATHIS WAGNER (Co-chair)

FABIO SCHIANTARELLI

This dissertation consists of three chapters.

The first chapter investigates how the historical development of Protestantism may contribute to explain current literacy disparities in India. Combining information about the spatial distribution of Protestant missions in India at the end of the nineteenth century with contemporary district-level data, I find a strong long-term relationship between the historical exposure to Protestant missions and current literacy. I then verify that this relationship is not driven by unobserved characteristics that may affect both current literacy outcomes as well as the missionaries' location decisions.

The second chapter exploits local variations in the historical exposure to Christianity to explain current differences in individual HIV-related sexual behaviors in Africa. I find that exposure to the presence of Catholic missions at the end of the nineteenth century is associated with a decrease in current HIV infection rates. I also examine whether historical Catholic and Protestant missions have a different impact on individual sexual behaviors. I find that Catholicism, while having a small negative impact on the propensity of condom use, is positively associated with the adoption of safer forms of sexual behavior (pre-marriage sexual abstinence, delay of first sexual intercourse and marital faithfulness).

Finally, in the third chapter I examine the impact of international migration and remittances on the labor supply of the family members left behind. Using data from Albania, I find that international migration has a significant impact on labor force participation. Remittances receipts from abroad determine a substitution effect away from the labor market, particularly for the female population.

Contents

Acknowledgements

I thank Scott Fulford, Mathis Wagner and Fabio Schiantarelli for their constant support and help. I am grateful to Samson Alva, Andrew Beauchamp, Matteo Cacciatore, Rossella Calvi, Donald Cox and Katherine Griwert 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, the Boston College Dissertation Workshop, the University of Waterloo and the University of California - Riverside.

Finally, I am deeply grateful to my parents and my siblings Luca and Silvia for their help, support and love.

1 The Protestant Legacy. Missions and Literacy in India.

1.1 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]

[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

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.

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-

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]).

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 impor-

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

tance 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 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

[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.

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.

1.2 Overview: Missionaries and Literacy in India

1.2.1 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 to public outrage in England,

[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)

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

[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,

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")

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*

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).

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

[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]).

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

1.2.2 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.

[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`

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]

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,

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

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

[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).

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.

1.3 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]

1.3.1 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

[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).

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 1 with the little crosses to indicate the location of 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 1 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

[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.

[20] Later in the paper I will use continuous variables in order to exploit also variations in the intensity of Protestant influence. See section 1.6.2 for more details.

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]

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

[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.

[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.

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 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

[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.

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.

1.3.2 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

[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]).

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 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.

1.3.3 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

[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.

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 1 is reproduced with the difference that, now, the districts have different colors according to how they perform in terms of literacy. The 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]

1.4 Empirical Approach

1.4.1 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:

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

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 spu-

[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).

rious 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 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.

1.4.2 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:

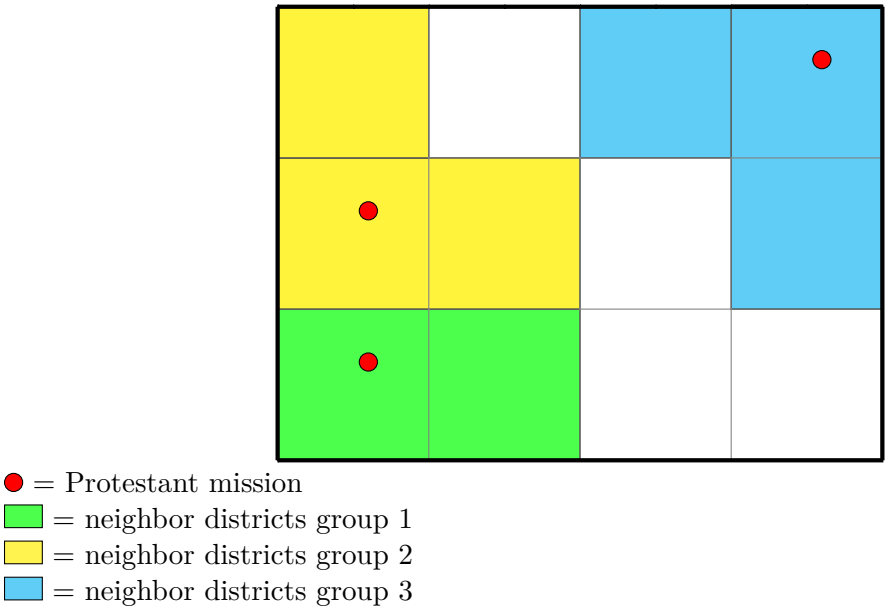
$$Y^{2001}_{i(g),j} = \alpha PM^{1908}_{i(g),j} + \mathbf{X_{i,j}}'\Gamma + \sum_{j=1}^{J}\sum_{g=1}^{G}\theta_{g,j}d_{g,j} + \epsilon_{i(g),j} \tag{3}$$

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.

1.4.3 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

Neighbor Districts



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 promo-

tion 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.

1.5 Results

1.5.1 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

[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.

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

[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.

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 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

[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).

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 alterna-

tive strategies.

1.5.2 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]

[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 10.

1.5.3 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 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]

[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).

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.

1.5.4 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]

[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).

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.

1.6 Robustness Checks

1.6.1 Outliers

Table 11 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 11 about here]

1.6.2 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.

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 12 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 12 about here]

[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 12 (results available upon request).

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 13 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 13 about here]

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 14 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

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

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 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 14 about here]

1.7 Potential Transmission Channels

The results presented in Section 1.5 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 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.

[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 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]).

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 1.12. 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,

[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

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.

1.8 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

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).

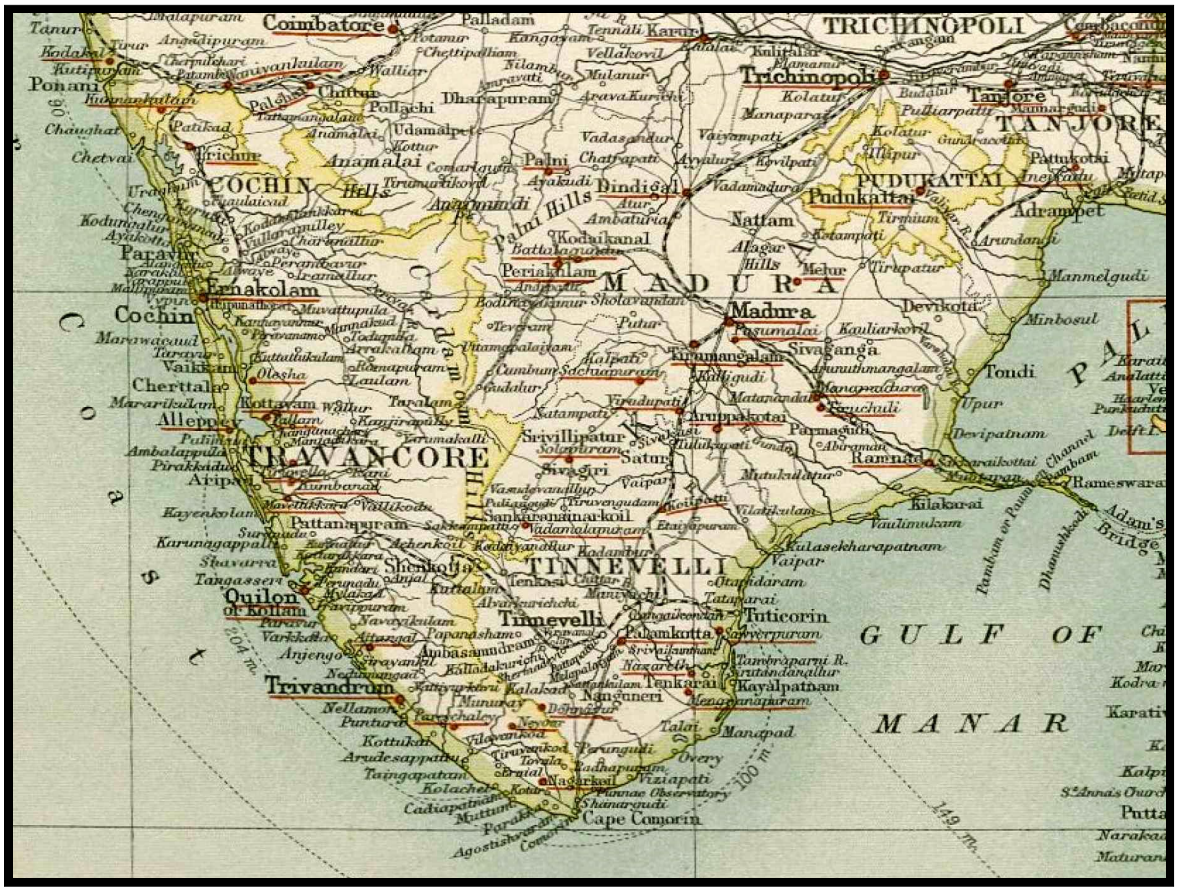
uncovered positive relationship. A placebo experiment using information about the location of Catholic missions around the same period of time lead to similar conclusions.

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.

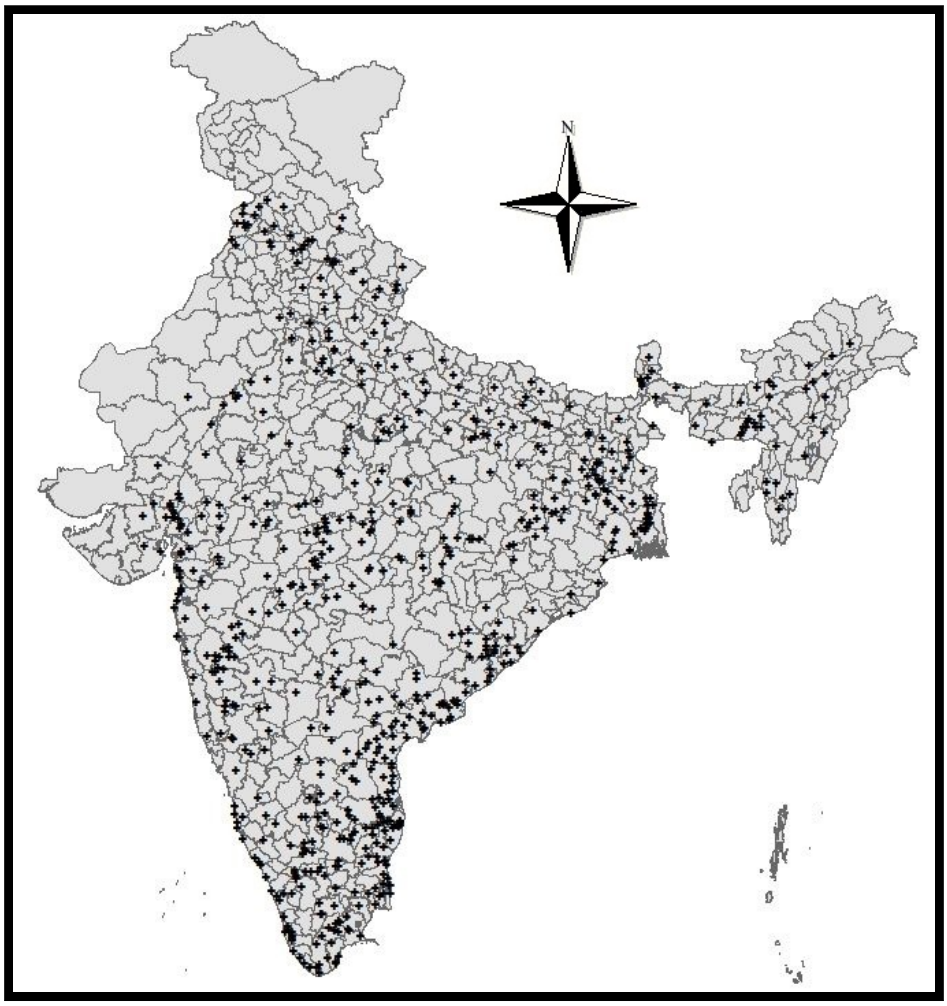
1.9 Figures

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



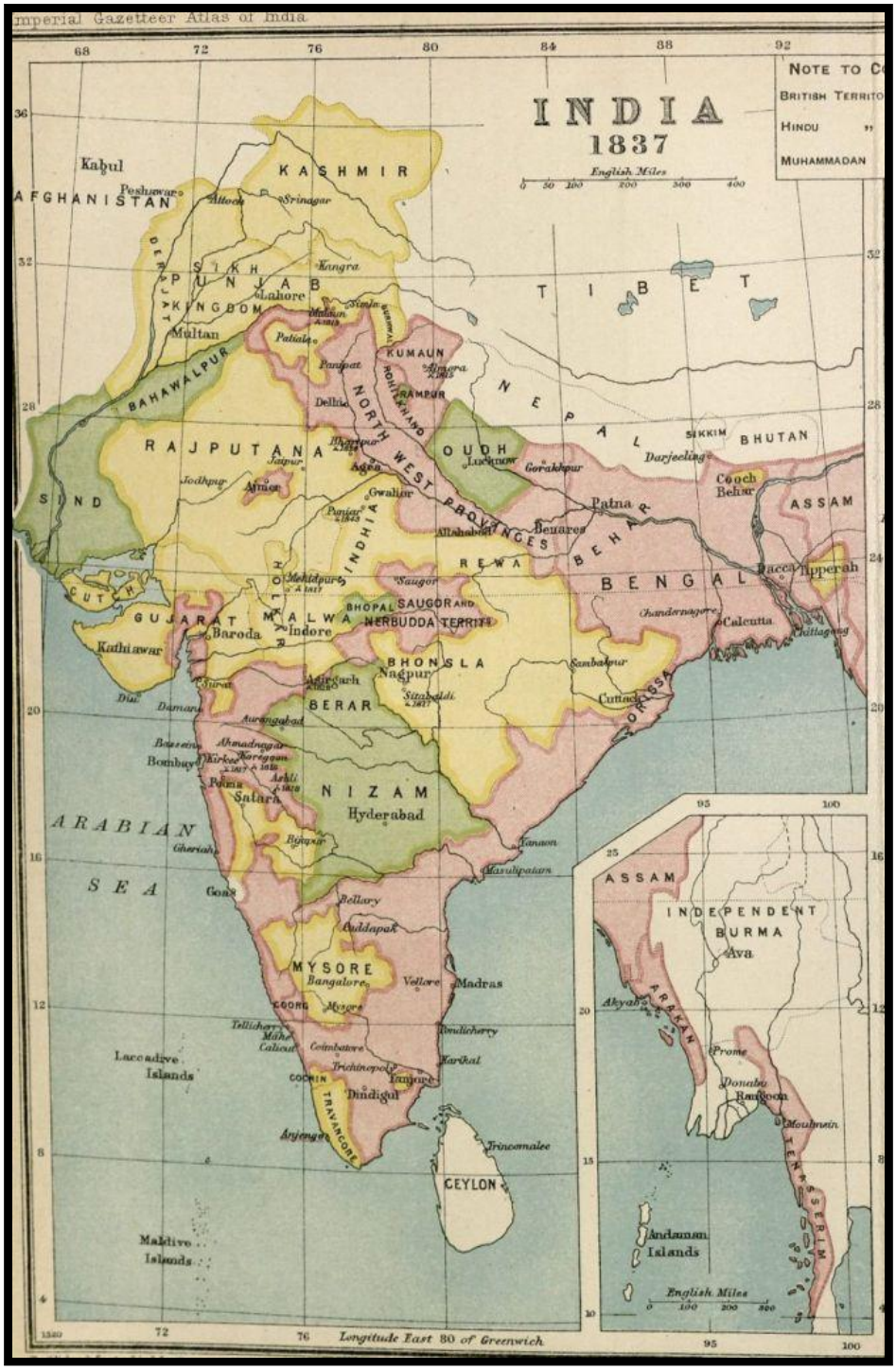
Note: towns with Protestant missions as of 1908 are underlined in red.

Figure 2: Historical Protestant Missions in India



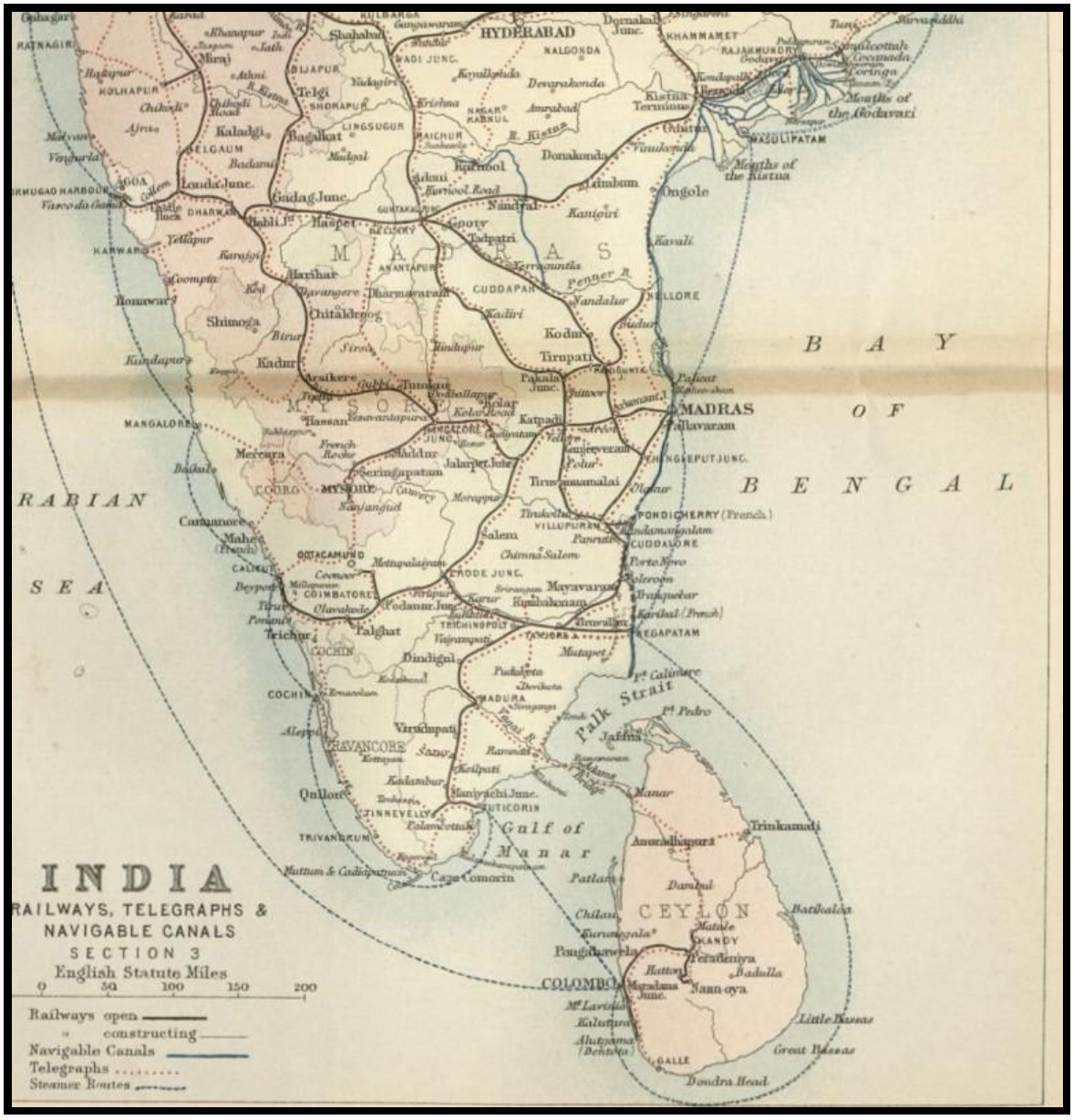
Note: this figure combines information about the location of Protestant missions in India as of 1908 contained in the "Statistical Atlas of Christian Missions" (little crosses) with a map showing the administrative boundaries of the Indian districts in 2001.

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.

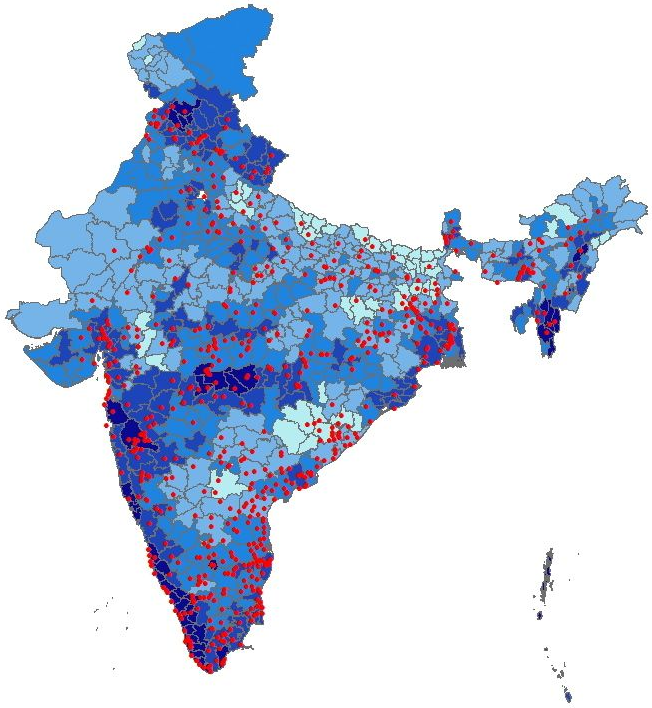
Figure 4: Railways in India as of 1891



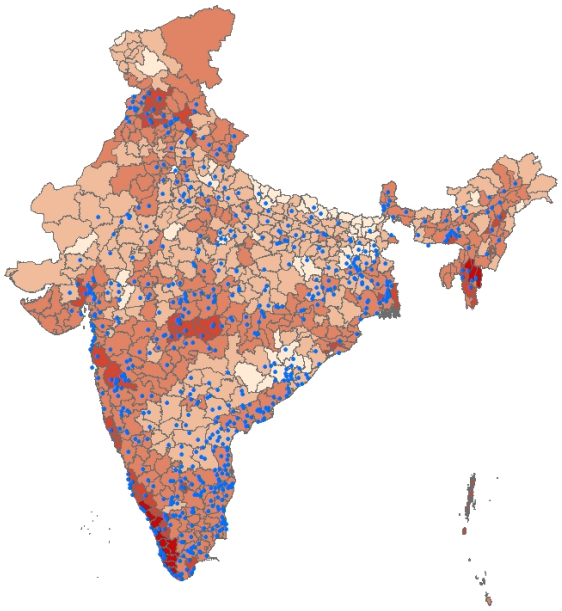
Note: section of Plate 20 from the "Constable's Hand Atlas of India" (1893) showing the railways network in India as of 1891.

Figure 5: Missions and Literacy Rates

(A) Population Literacy Rate



(B) Female Literacy Rate



(C) Gender Parity Index

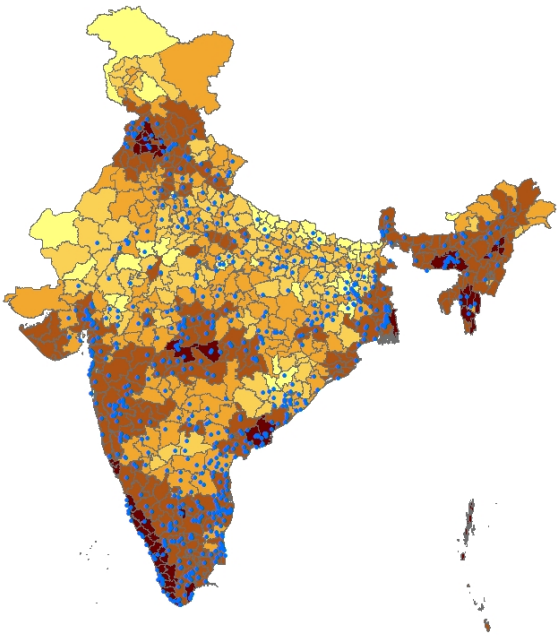
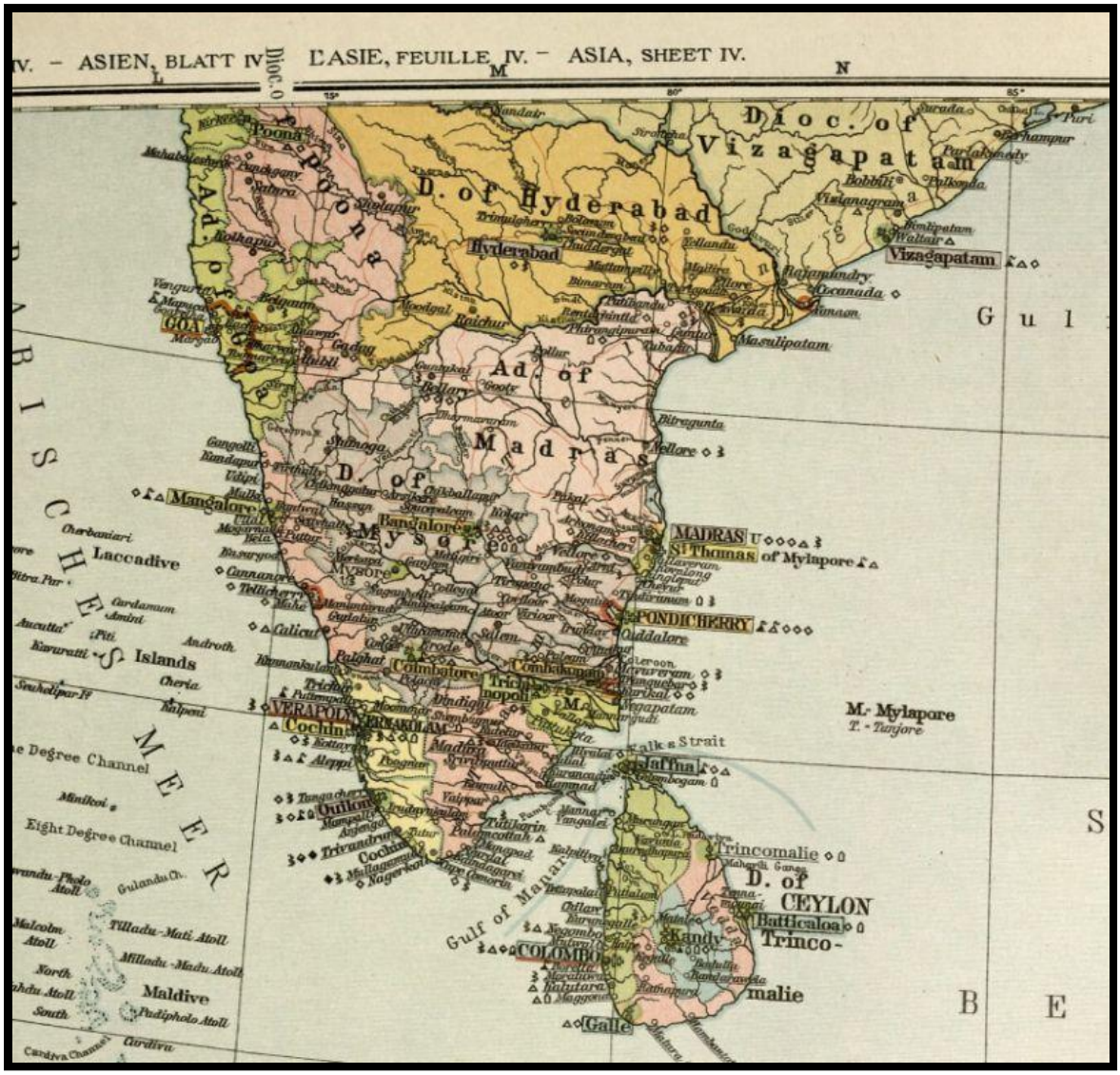
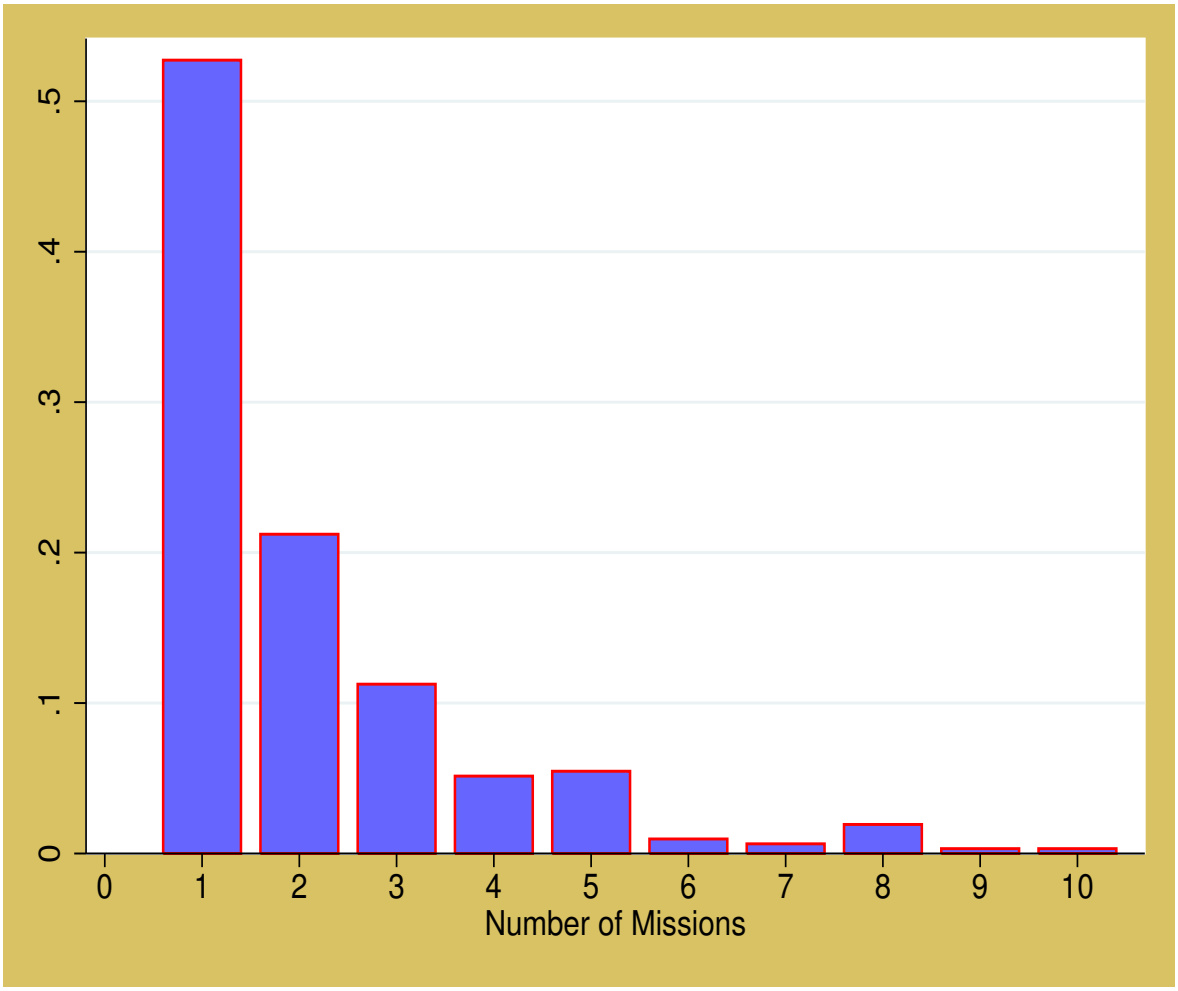


Figure 6: Map's detail from the "Atlas Hierarchicus" (1913).



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.

1.10 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 the ratio between female and male literacy rates. 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.Dens. 1901	No Agric. 1901	Muslims 1901	Infirm. 1901	Rails 1891	Tot. Lit. 1901	Female Lit. 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	Female Lit. Rate	GPI
	[1]	[2]	[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

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 1.12 for more details.

1.11 Appendix A: Additional Tables

Table 10: 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 10 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 11: 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	

Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 12: Using Historical Missions Density as Measure of Missionary Influence

	Full Sample				N.D. Sample
	[1]	[2]	[3]	[4]	[5]
Dep. Variable: Total Population Lit. 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

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 13: 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

Robust standard errors, corrected for clustering within states, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent level.

Table 14: 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. Dist. from	-0.096***	-0.093***	-0.036**	-0.127***	-0.174***	-0.178***	-0.094**	-0.252***	-0.236**	-0.301**	-0.194**	-0.291***
Prot. Miss. (/100)	(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. Dist. from	0.022	0.021	0.004	0.025	0.002	-0.005	-0.015	-0.002	-0.004	-0.132	-0.075	-0.136
Cath. Miss. (/100)	(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 14 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.

1.12 Appendix 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 $\Delta\widehat{\alpha}$.

2 Christian Missions, HIV and Sexual Behavior in Sub-Saharan Africa.

2.1 Introduction

Existing studies have pointed to a host of factors to explain the epidemiology of the HIV pandemic.[41] In correspondence with the larger involvement of faith-based organizations in HIV/AIDS prevention and support activities, researchers have started focusing on the role of religion and religious organizations.[42] By shaping family and community values around specific HIV-related sexual behaviors, religious groups can be instrumental in reducing or contributing to the spread of the virus. The extent to which religious institutions affect individual values and, consequently, behaviors depends on the organizations' influence over the local cultural environment. Consistent with that view, in this paper I explain current differences in HIV infection rates and individual HIV-related sexual behaviors in Sub-Saharan Africa by exploiting geographical variations in the exposure to Christianity at the beginning of the twentieth century.[43] Preferences, values and social norms are transmitted across generations and acquired via different forms of social interactions as the result of a cultural transmission process (Bisin and Verdier [2000]). A community that has been exposed to Christianity over several decades is therefore likely to have developed specific values

[41] Attention has been devoted, in particular, to medical (such as the coexistence of other sexually transmitted disease), behavioral (pre- and post-marriage sexual behavior) and cultural (social position of women, circumcision, poverty, migratory pressures) factors. For an overview of the existing research, please refer to Caldwell (2000),Brockerhoff and Biddlecom (1999), Oppong (1995), Kohler et al. (2007), Bongaarts (2007), Reniers (2008).

[42] In Namibia, for example, Catholic AIDS Action, established in 1999, operates throughout the country offering home-based care and counseling to approximately 3,000 households (Yates [2003]). A recent survey conducted in Zambia (ARHAP [2006]) shows that the large majority (80 percent) of religious support groups offers an HIV/AIDS service.

[43] The choice of focusing on Sub-Saharan Africa is due to the fact that the region has been heavily affected by the HIV pandemic. It accounts for more than two-thirds of all people living with HIV and nearly three quarters of AIDS-related deaths (UNAIDS [2010]). Data also indicate a reversal in life expectancy due to AIDS mortality. In southern Africa, average life expectancy at birth is estimated to have declined to levels last seen in the 1950s; it is now below 50 years for the subregion as a whole and below 40 years in Zimbabwe (WHO [2006]).

that persisted over time and affect today's individuals' sexual conduct.

Recent research has offered statistical evidence showing that historical events can have long-term impacts that continue to be felt today. Increasingly, attention has turned to better understanding the specific mechanisms underlying historical persistence.[44] Among the channels through which history affects current socio-economic outcomes, culture plays a potentially important role because it represents a slow moving variable whose evolution can be affected by historical events. Most famously, the idea that religious history, through its impact on values and culture, may persistently affect current outcomes originates from Max Weber's hypothesis that the Protestant Reformation was instrumental in facilitating the rise of industrial capitalism in Western Europe. Although there is no empirical consensus around Weber's claim,[45] attempts to test its validity led to an enrichment in our understanding of the mechanisms through which historic religious events may contribute to current socio-economic development. Recent papers find a positive long-term effect of Christianity (Protestantism, in particular) on human capital and knowledge diffusion. Becker and Woessmann (2009) show that in ninetieth century Prussia, Protestant counties displayed higher literacy rates than counties under Catholic influence. Using information about the historic presence of Protestant missionaries, Gallego and Woodberry (2010) provide evidence of a long-term positive impact of Protestantism on educational outcomes in a set of African provinces.

Building on that research, I examine whether Christianity has a long-lasting effect on HIV/AIDS-related sexual behaviors. To this aim I use the geographical variation in the exposure to Christian culture during the colonial period (end of ninetieth

[44]The seminal works are Engerman and Sokoloff (1997), La Porta et al. (1997) and Acemoglu et al. (2001). For an analysis of the effect of history through institutional development see Djankov et al. (2003), Acemoglu et al. (2005). The impact of history on the diffusion of knowledge and technology is the focus of the analysis in Glaeser et al. (2004), Comin et al. (2010). Finally, Tabellini (2007), Nunn and Wantchekon (2011) look at how history affects levels of social capital and trust.

[45]For example, Iannacone (1998)'s survey concludes that empirical evidence tends to reject Weber's thesis. More recently Cantoni (2010) finds no evidence of the effect of Protestantism on economic growth in Germany.

century-first decade of twentieth century) as a quasi-natural experiment. To measure the degree of long-term exposure to Christianity I rely on information contained in two historical atlases. The *Statistical Atlas of Christian Missions* provides maps regarding the worldwide location of Protestant missions as of 1908. The *Atlas Hierarchicus*, an official Vatican document, contains detailed charts with information about the location of Catholic missions as of 1911. Using geo-referencing techniques I combine this information with the location of residence of today's individuals and I approximate the degree of historic contact with Christianity as the local number of Protestant and Catholic missions during the colonial period. I then combine this variable with data from the Demographic Health Surveys which provide information regarding individual HIV status and sexual behaviors for a sample of sub-Saharan countries characterized by high HIV infection rates. I have made an effort in order to control for the factors that determined the location of Christian missions during the colonial period and that may be potentially correlated with today's individuals' sexual behavior, too. Historiography indicates geography as a crucial determinant of the missionaries' location decisions. Accordingly I control for geographical differences across villages as well as for country and regional fixed effects. Moreover, early missionaries tended to locate in proximity of commercial routes in order to maintain access to supplies from Europe. To take this aspect into account, I rely on historic sources containing information about the main exploration routes in the African continent during the ninetieth century.

The results of the paper show a long-term negative effect of Catholic missions on the HIV infection rate in the region. On average, an additional Catholic mission operating in a village during the colonial period determines a 1.2 percentage points decrease in the infection rate of its current residents. Protestant missions, on the other hand, do not appear to have any significant impact. Given that sex represents the main vector of HIV transmission in the area, this finding may be indicative of the

fact that the two religious denominations have different effects on individual sexual behaviors. I focus, in particular, on behaviors closely related to the pillars of the so-called "ABC strategy" against HIV, a sex education policy, common in several African countries, that emphasizes: (A) sexual abstinence for youth and unmarried individuals; (B) mutual faithfulness and partner reduction for sexually active adults; (C) a correct and consistent condom use.

The Catholic and Protestant denominations have different views regarding the use of condoms. It is well known that the Catholic Church does not accept any form of sexual contraception because it renders procreation (the only purpose of sexual activity) impossible. This ban extends to the use of condoms, even in cases where such use may help in preventing the spread of a sexually transmitted disease.[46] Accordingly, I find that living in a community that has been historically exposed to the presence of a Catholic mission negatively affects the likelihood of condom use at last sexual intercourse. Protestant missions, on the other hand, do not exert any significant impact. This may be explained by the fact that the position of the traditional Protestant churches regarding contraception methods has historically been less conservative. Since the 1950s, for example, the Anglican Church has stopped considering the use of forms of contraception as a sin or a contravention of God's teachings. And, in response to the spread of AIDS/HIV in the poorest regions of the world, many Protestant denominations have recently backed the use of condoms as an additional way to limit the pandemic.

I find evidence of a differential effect of the two religious denominations also as concerns the other two types of sexual behavior considered here. While Catholic missions increase the individual likelihood of having abstained from sex before marriage (that is the period of more intense sexual promiscuity and, potentially, of higher risk of infection), living in a village that has seen the historical presence of Protestant

[46]Still in 2009, Pope Benedict XVI stated that the AIDS pandemic is a tragedy that "cannot be overcome through the distribution of condoms".

missions is associated with a decrease in that same probability, particularly for males living in urban areas.

Christian teachings consider adultery immoral. Given the strong condemnation of marital unfaithfulness by Christianity, I expect a negative impact of Christian missions on this type of sexual conduct. The findings support this expectation: an additional mission in the village determines a 0.2 percentage points decrease of today's individuals' propensity of engaging in marital unfaithfulness. When I distinguish between the two religious denominations, I find that this effect is largely accounted for by the presence of Catholic missions: a one unit increase in the village's number of historic Catholic stations is responsible for a 0.8 percentage points decrease in the occurrence of extra-marital sexual relationships.

The result that historic Catholic missions are related to a lower HIV infection rate while decreasing the individual propensity of condom use is in line with some recent findings in the medical literature. Hearst and Chen (2004), Shelton (2007) and Potts et al. (2008), for example show how, in the case of sub-Saharan Africa, there is no evidence of a primary role of condom use in generating a measurable slowing of new infections; most of the success in the fight against AIDS in the region has to be attributed to the promotion of safer preventive forms of sexual behaviors such as a reduction of the number of sexual partners.

In the last part of the paper I look at three channels through which Christian missions may have operated in order to affect current individual behaviors: education, religiosity and culture. The results show how historic Christian missions maintain a significant long-term effect on sexual behaviors even when controlling for the individual level of human capital and religious belief. One reason may be that behaviors (and sexual conducts in particular) are largely influenced by peer effects and other types of social pressures. It can be for example that, conditioning on the same level of education and religious faith, an individual living in a relatively more liberal society

may adopt behaviors that are completely different from those adopted by someone living in a more conservative community. Taking these social pressures into account significantly reduces the impact of Protestant and Catholic missions on all the three types of sexual behaviors. I interpret this result as suggestive of the fact that most of the long-term impact of historical missions on individual sexual behavior is due to the development and persistence across generations of a set of moral norms that shaped the cultural environment at village/community level leading its residents to adopt (or to avoid) specific behaviors as the result of social influence and peer effects.

The remainder of the paper is organized as follows. Section 2 offers an historical background, briefly describing the development of Christianity and Christian missions in Sub-Saharan Africa. Section 3 describes the data sources and the main variables used in the empirical analysis. Section 4 presents the results concerning the long-term effect of Christian missions on HIV infection rates and HIV-related sexual behaviors. Section 5 analyzes the alternative channels through which Christianity persistently affects individual behaviors. Finally, section 6 concludes.

2.2 Historical Background

Contemporary Christianity in Africa dates back to the ninetieth century.[47] The evangelization of the continent (particularly the sub-Saharan region) took place in a relatively short period of time and embraced an extremely large sector of the population so that it is considered the greatest epoch of conversion in the entire history of Christianity. Two factors can be adduced for this phenomenal development. On one side the religious revival in Europe led to a great global missionary effort. On the other side the almost contemporaneous occurrence of the exploration of the continent opened up Africa to the influence of the Western culture.

As a starting point of the whole epoch the year 1792 can be mentioned as it saw

[47]This section is based, primarily, on Baur (1994) and Robinson (1915).

the foundation of the first Protestant missionary society to work in Africa (Baptists), the establishment of the first mission in South Africa (Moravian Genadental) and the foundation of Freetown in Sierra Leone which will become the main outpost of Protestant mission work in West Africa. Initially evangelization was, with few exceptions, restricted to liberated slaves only. Proper missionary activity started in 1850s and the period between 1880 and World War I (1914) is referred to as the most decisive period in African Christian history.

The ingredient that fostered modern Christianity in Africa is Protestantism. For half a century (1792-1850) Protestant missionaries were practically alone in the field and, at least until the first decades of the twentieth century, Protestant missions were far more numerous than Catholic ones. The major reason for this is that in the previous century the various Protestant denominations in Europe had experienced a great revival movement, while the Catholic Church had suffered a serious decline started with the Napoleonic Wars (1815 circa). From the area around the Cape Town colony, several Protestant groups (such as the Pietists and the Moravians from Germany and the Methodists and the Evangelical wing of the Anglican Church from England) began to diffuse their presence and activity in all the southern region of Africa. One important factor that fueled modern evangelization was represented by the extensive geographical explorations in the ninetieth century. Around 1800 Africa still remained the only continent whose interior was basically unknown (hence the name "the dark continent"). This naturally attracted the interest of explorers who, more or less intentionally, opened new paths for the missionaries. David Livingstone (a former missionary of the London Missionary Society himself), for example, between 1852 and 1856 opened a route from the Upper Zambezi (today's Zambia) to Central and Eastern Africa.[48] Following Livingstone's footsteps, missionary societies started

[48]David Livingstone (1813-1873), a member of the *London Missionary Society*, dedicated his life to opening exploration routes in central and southern Africa, with the goal of preparing the region to Christian evangelization. At the historic meeting held in the Senate House in Cambridge (England) in 1857, Dr. Livingstone made his goal clear by ending his address with the words: "*I know that in*

to establish their stations in the area between Zambia, Mozambique and Kenya.[49]

Catholic missions followed the Protestant ones a full generation later, in correspondence with the papacy of Gregory XVI (1831-1846) who started a reorganization of the missionary activity of the Catholic Church attributing its supervision to the Congregation for the Propagation of the Faith. This was followed, in the second half of the ninetieth century, by the revival and missionary re-orientation of old orders (Benedictines, Dominicans, Franciscans and Jesuits) as well as by the foundation of new congregations which dedicated themselves specifically to the work of evangelization. During the whole ninetieth century South Africa was almost exclusively under the influence of Protestantism, with Catholics remaining an insignificant minority until the second half of the twentieth century.[50] Given the Protestant preponderance in the South, Catholics looked at the Eastern African region as their preferred mission field. The first missions were established by the Capuchins (in the Ethiopian territory of the Omoro tribe in 1846) and by the *Holy Ghost Fathers* (a French Catholic denomination) on the island of Zanzibar in 1860 and, later in 1868, in Bagamoyo on the Tanzanian coast. From there, Catholics began their work of evangelization in the area covering Tanzania, Kenya and Uganda (in 1878 ten missionaries from the *White Fathers* congregation established outposts along the coasts of Lake Victoria and Lake Tanganyika), sometimes in strict competition with the Protestant societies that, by late 1880s, also started to establish their missions in the region (particularly in the

a few years I shall be cut off in that country which is now open: do no let it be shut again. I go back to Africa to try and make an open path for commerce and Christianity [...]" (Robinson [1915], page 318).

[49]In Zambia the first settled Christian missions were the *Paris Mission* (1877) and the *London Missionary Society* (1883), followed by the *Presbyterians* (1894), the *Primitive Methodists* (1894), the *White Fathers* (1895), and the *Dutch Reformed Mission* (1899). The evangelization in Kenya started in 1844 with the arrival of the *Anglican Church Missionary Society*; other European missions were soon to follow, including *British Methodists* (1862), the *Scottish Presbyterians* (1891) and the interdenominational *Africa Inland Mission* (1895).

[50]The only significant exceptions are represented by the French *Oblates of Mary Immaculate*'s mission in Lesotho and the *Mariannhill* mission, founded in 1882 near Durban by Austrian Trappists

area around Lake Victoria).[51]

2.3 Data

The empirical analysis depends on the availability of two types of data. On one hand I need data about the sexual behaviors and HIV status of individuals living today in sub-Saharan countries. On the other hand, I need information about the location of Christian missions in the region during the ninetieth century.

The source for individual and household data is represented by the Demographic and Household Surveys (DHS, hereinafter); historical data concerning the location of Christian missionary stations come from an original dataset built specifically for this research.

2.3.1 Historical Christian Missions

I retrieve information about the location of the Christian missionary stations at the beginning of the twentieth century from two historical atlases.

With regard to the location of Protestant missions, I rely on the *Statistical Atlas of Christian Missions*, produced in occasion of the World Missionary Conference of Edinburgh (Scotland) in 1910. The *Atlas* represents a directory of missionary societies throughout the world and a statistical presentation of their operations in each country 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 it includes only those missions that conduct their operations among non-Christians. Therefore, only the Protestant missions that operate among indigenous populations are recorded; missions meant to serve immigrants (i.e. Westerners) are not included in the data. The *Atlas* reports very detailed statistics

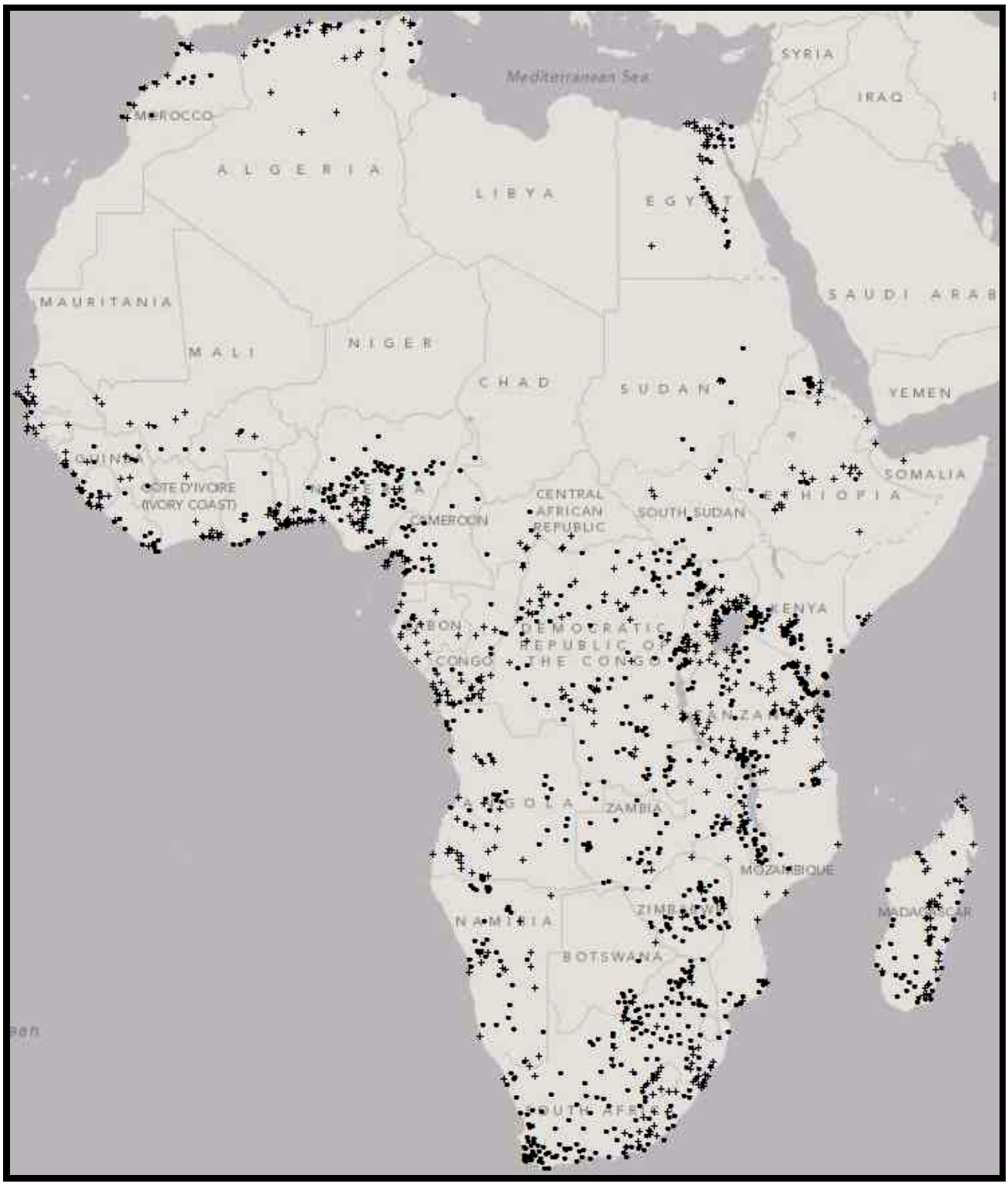
[51]The coexistence of Catholics and Protestants in the area was not without episodes of tension. The rivalry reached, dramatically, its peak in 1892 with the *Battle of Mengo*. Order was then re-established in 1900 with the *Uganda Agreement*, arranged after the intervention of the British government, with a clear division of Ugandan counties between Catholics and Protestants.

about the worldwide operations and the organization of each Protestant Society. So, for example, it is possible to know that, as of 1908, the *United Free Church of Scotland* had 28 missionary stations in South Africa, for a total of 123 missionaries. Alongside with this large amount of statistics, the *Atlas* offers a series of maps displaying the geographical location of the missions (an example of the type of maps contained in the *Atlas* is presented in Figure 3 in the Appendix). Using cartography tools, I superimpose those charts to a digital map of the African continent obtaining what can be seen in Figure 1, with the • symbol indicating the location of the main Protestant missionary stations as of 1908.

The *Atlas Hierarchicus* (AH) elaborated by Karl Streit in 1913 is the source used to identify the villages and cities with a Catholic mission. The AH, an Vatican official document, meant to represent a survey of the global hierarchy of the Catholic Church as also a delineation of its territorial division. The AH provides very detailed maps that identify all the single ecclesiastical districts (dioceses) and the Catholic institutions operating in each of them: episcopal towns (distinguished according to their rank), the general and provincial residences of the male orders and congregations, abbeys, monasteries and places of pilgrimage. Relevant to this research, the charts report also the geographical location of the Catholic missions as of 1911 (for an example of the type of maps contained in the AH, see Figure 4 in the Appendix). Using the same procedure as before I was able to geo-reference the location of each mission established in Africa as of 1911 and I added them to the map in Figure 1, where the Catholic missions are now indicated by a + symbol.

Looking at the map, it is possible to notice that, although there is a near clustering of Protestant and Catholic missions, some amount of mixing occurs as many areas see the presence of missions of both types. This mixing is likely to be the direct consequence of the freedom of religion that had been guaranteed throughout Africa

Figure 1: Location of Catholic (+) and Protestant (•) missions in 1908-1911



Source: author's elaboration from the "*Statistical Atlas of Christian Missions*" (1910) and the "*Atlas Hierarchicus*" (1913).

by the General Act of the Conference of Berlin (1885).[52] Protestant missions represent the large majority of the missionary stations in the area south of the Zambezi river (Zambia, Zimbabwe, Malawi and South Africa) while the Catholics are more concentrated in Eastern Africa and in the Great Lakes region (Kenya, Uganda and Tanzania).

I then combine the location of Christian missions with the geographical location of residence of the individuals surveyed in DHS.[53] Around each village I draw a circle of a radius of 30 kilometers and I simply record the number of missions (total Christian, Protestant and Catholic) contained in it.[54] This count represents my measure of the degree of historic local exposure to Christianity.

The top panel of Table 3 shows that 46 percent of the sampled individuals live in a village that had at least a Christian mission by 1911. Some degree of variability is observed among countries, with Kenya displaying the largest concentration of missionary societies and Zambia the lowest. The data also confirm the coexistence of both Catholic and Protestant missions in the same location. On average, 14 percent of the respondents reside in a village that saw the contemporaneous presence, during the colonial period, of missions belonging to both religious groups (this is especially the case of the Democratic Republic of Congo, Kenya and Zimbabwe). Finally, Catholic missions appear to have the tendency of clustering near each other. Conditional on the number of missions being positive, on average a village sees the presence of 2.9 historical Catholic missions (with Kenya displaying the largest concentration),

[52]"All the Powers exercising sovereign rights or influence in the aforesaid territories [...] shall, without distinction of creed or nation, protect and favour all religious, scientific or charitable institutions[...]. Christian missionaries, scientists and explorers, with their followers, property and collections, shall likewise be the objects of especial protection. Freedom of conscience and religious toleration are expressly guaranteed [...]. The free and public exercise of all forms of divine worship, and the right to build edifices for religious purposes, and to organize religious missions belonging to all creeds, shall not be limited." (General Act of the Conference of Berlin, Chapter I, article 6)

[53]For confidentiality reasons, in DHS up to 2 kilometers of random error in any direction is added to villages located in urban areas, which becomes up to 5 kilometers for those located in rural areas.

[54]The choice of a 30 kilometers radius is arbitrary. No significant changes occur when using a smaller (10 km) or larger (50 km) radius.

compared to 1.8 historical Protestant missions.

2.3.2 Individual and Household Level Data

Individual data are from the Demographic Health Surveys.[55]DHS data are collected in several countries across the world using a standardized sampling design and common questionnaires that allow cross-country comparisons in topics related to health care (specifically maternal and child care).

In the analysis I focus on Sub-Saharan Africa. Countries in that region have been particularly exposed to the spread of HIV/AIDS. It is therefore interesting to understand what are the factors that affect the preventive and protective sexual behavior of the individuals living in that area.

Not all the DHS surveys contain geocoding information; since this is crucial for the empirical strategy of the paper, its lack necessarily limits the number of countries that can be included in the study. Table 1 reports the list of countries used in the empirical analysis, the year of the survey, and the corresponding HIV prevalence rate.

I restrict the sample to individuals that are classified as usual residents of the household (who represent 97.01% of the total sample) since the DHS do not provide information about the place where the visitors live and it is therefore impossible to link them to the Christian missions' dataset. At the end, the final sample consists of 42,355 individuals, belonging to 2,567 villages.

Table 1: **Sample of Countries**

Country	Survey	HIV Prevalence Rate
Democratic Republic of Congo	Standard DHS, 2007	4.5%
Kenya	Standard DHS, 2009	6.7%
Malawi	Standard DHS, 2010	11.9%

[55]`http://www.measuredhs.com/`

Tanzania	Standard DHS, 2003	6.2%
Zambia	Standard DHS, 2007	15.2%
Zimbabwe	Standard DHS, 2006	15.3%

Note: data for the HIV prevalence rate are from the *CIA World Factbook* (2011).

Outcome Variables

Since 2001, in an effort of better understanding the HIV pandemic, the DHS started releasing information regarding the results of an HIV test conducted in a voluntary and anonymous form. The testing's procedure consists of the collection of blood drops from a finger prick on a filter paper. The blood spots are then dried and transported to a laboratory for testing.[56] DHS provide HIV testing results for all the countries in my sample. In particular, for almost the totality of the sampled individuals (42,328 over 42,355) I am able to build a binary variable taking value equal to one if the HIV test is positive and equal to zero otherwise. Table 2 shows how the HIV prevalence rate is slightly higher for the female sample. This is in line with most of the empirical evidence regarding the spread of HIV in Sub-Saharan Africa (UNAIDS [2007]).

Transmission of HIV through sexual contact has been the most frequent means of the spread of the disease. Understanding the determinants of individual sexual behavior (i.e. what leads people to adopt certain types of sexual behavior) is crucial in order to isolate the factors that can limit the diffusion of the virus. To this aim, the DHS provide a wide range of information about sexual behaviors and attitudes. I focus on measures that represent the three pillars of the so called "ABC strategy" against HIV. Starting from the 1990s, the ABC approach has been developed in response to the growing epidemic of HIV/AIDS in Africa and, more in general, to prevent the spread of sexually transmitted diseases. This sex education policy employs population-specific interventions which emphasize: (**A**) abstinence for youth

[56] In the laboratory the sample undergoes an initial ELISA (Enzyme-linked Immunosorbent Assay) screening. In a second stage, all the positive samples and 10 percent of the negative samples are retested with a second ELISA. For those tests with discordant results on the two ELISA screenings, another test, usually a Western blot, is used to determine the final result.

and other unmarried people; (**B**) mutual faithfulness and partner reduction for sexually active adults; (**C**) a correct and consistent use of condoms.

For each sexually active individual, the DHS report the respondent's age at sexual debut and the age when she/he got married for the first time. Based on this information I create a binary variable equal to one if the individual has waited until marriage to have his/her first sexual intercourse, and zero otherwise.

The risk of getting infected increases with the frequency of sexual encounters and, more importantly, with the identity of the sexual partner. It is likely that individuals who have sex only with their spouse are less likely to be infected relative to those that have casual sexual encounters. Through a set of retrospective questions, the DHS provide an insight of the respondent's sexual behavior in the year prior to the interview. In particular, individuals who are currently married (or who are living with their partner) are asked whether they had any sexual experience with anyone else beyond their spouse/partner in the last twelve months. This allows me to create a binary variable which takes value equal to one if the respondent did have an extramarital sexual experience and zero otherwise.

The "C" component of the ABC approach advocates the correct and consistent use of condoms during sexual intercourse as the third pillar of an effective strategy to limit the spread of the HIV pandemic. While it is not possible to derive a perfect measure of a consistent condom use from DHS, the surveys report whether a condom was used during the last sexual intercourse. I therefore build a dummy variable equal to one if that was the case, and zero otherwise. Although not perfect, I use such variable as a proxy for the third pillar of the ABC strategy.

Table 2: Outcome Variables

Variable	**Mean**	**S.D.**	**Obs**
		Males	
HIV-Positive	0.10	0.30	19,403

Condom Used	0.19	0.38	19,403
Extra Marital Sex	0.13	0.33	14,911
Abstinence	0.09	0.28	13,143
		Females	
HIV-Positive	0.12	0.33	22,790
Condom Used	0.08	0.28	22,790
Extra Marital Sex	0.03	0.13	19,731
Abstinence	0.21	0.40	11,792

Table 2 reports the descriptive statistics for the outcome variables used in the empirical analysis.[57] Some specific patterns emerge. Men result to be more likely than women to have used a condom during their last sexual intercourse; however they also display a riskier sexual behavior as they tend to be more likely than women to have experienced extra-marital sex. In addition, on average women tend to display a higher probability of delaying their first sexual experience until after marriage, relative to men.

Control Variables

In the empirical analysis I include several individual and household characteristics. Alongside the usual demographic information (age, age squared, a gender indicator), I also control for the marital status of the individuals and their employment status in the last twelve months.[58] The DHS do not contain income information. In order to take into account the household's level of wealth, I include a set of dummy variables indicating the household ownership of a series of assets (electricity in the house, radio, television, bicycle, type of roof and type of floor). As an alternative, the surveys provide a wealth index (built via Principal Component Analysis) that divides the

[57]Notice that the number of observations is larger when the outcome variable is *Condom Use*. This is due to the fact that when considering *Extra-Marital Sex*, by construction only the sample of currently married individuals is used; when, instead, the outcome variable is *Abstinence*, the sample consists of respondents who are currently married or who were formerly married.

[58]Unless otherwise specified, I consider as *married* both those who are actually married and those who live with their partner.

sampled households into 5 wealth categories.[59] Using the wealth index instead of the individual asset dummies does not affect the results.

Sexual behavior may also depend on how much an individual knows about HIV and the ways the virus spreads. It is likely that better informed people are also more inclined to adopt safer forms of sexual behavior. For this reason, I control for the individual level of knowledge about HIV/AIDS via the inclusion of an index based on a set of six questions in the DHS.[60] For each question, I observe whether the individual answers correctly, wrongly or if she/he does not know the answer. Each case reveals a given type of information about HIV/AIDS knowledge. The main concern is how to deal with the "don't know" answer. In the empirical literature using survey data, these cases are often recoded as missing values. In this context, though, the "don't know" answer reveals more information than a missing answer because it is indicative of ignorance about safe and risky sexual practices. To take that into consideration, for each question I assign a score equal to 1 in case of a correct answer, to -1 in case of an incorrect answer, and to zero in case the individual declares not to know the answer. Summing all the single scores, I end up with an index of "AIDS-knowledge" ranging from -6 (worst knowledge) to +6 (best knowledge).

The degree of exposure to the AIDS epidemic may influence how individuals behave, too. Knowing someone who is HIV-positive may increase the likelihood of adopting safer forms of sexual behavior due to the fear of becoming infected too. To take that into account, I create a binary variable equal to one if the agent knows someone who has HIV or died because of AIDS, and zero otherwise. Likewise, having been tested for AIDS in the past can be a predictor of individual sexual behavior

[59]The wealth categories are: poorest, poorer, middle, richer, richest.

[60]The questions are: 1) *Can people reduce their chances of getting the AIDS virus by using a condom every time they have sex?*; 2) *Can people reduce their chances of getting the AIDS virus by having just one partner who is not infected and who has no other partners?*; 3) *Can people reduce their chances of getting the AIDS virus by not having sex at all?*; 4) *Is it possible for a healthy-looking person to have the AIDS virus?*; 5) *Can a person get the AIDS virus from mosquito bites?*; 6) *Can people get the AIDS virus by sharing food with a person who has AIDS?*

since this implies that the respondent has received some form of pre- and post-test counseling, which represents the most customized way of transmitting information about AIDS and preventive methods. I control for this aspect by including a dummy variable equal to one if the individual has been tested at least once prior to the survey, and zero otherwise.

These variables represent the core of the control set in the empirical analysis. The middle panel of Table 3 reports the related descriptive statistics for the entire sample and country-by-country. On average there is a slightly larger representation of the female group, mainly due to the sampling design occurring in Tanzania and Zimbabwe that led to a female over-representation. The rate of participation in the labor market displays some level of heterogeneity at country level with Zambia and Zimbabwe under-performing relative to the other countries in the area.

Looking at AIDS/HIV-related variables, we can see how they are quite heterogenous across countries. Even if almost the totality of interviewed people (around 98 percent) declares to have heard of AIDS, the actual knowledge of the disease, measured combining the answers to six direct questions in DHS, is far from being perfect. Respondents in Kenya and Tanzania seem to have a better knowledge of the virus and the ways through which it spreads, particularly in comparison to respondents in Congo DR who show a level of knowledge significantly below the sample mean. Finally, almost half of the respondents declares to know someone that has been infected by the virus or has died because of AIDS and almost the same proportion of individuals has been tested for AIDS at least once, with Malawi reporting the highest test rate among the sampled countries.

2.3.3 Village Level Data

Christian missionaries did not locate randomly in Africa. Historiography (see Johnson [1967] and Baur [1994]) identifies a number of factors as determinants of the location

Table 3: Descriptive Statistics

	All	Congo DR	Kenya	Malawi	Tanzania	Zambia	Zimbabwe
Historical Missions							
Pr(ChrisMiss = 1)	0.46	0.55	0.66	0.43	0.53	0.34	0.49
	(0.49)	(0.49)	(0.47)	(0.49)	(0.49)	(0.47)	(0.49)
# ChrisMiss	2.20	2.13	3.63	1.69	1.83	1.44	2.34
	(1.49)	(1.30)	(2.00)	(0.98)	(1.35)	(0.93)	(1.17)
Pr(only CathMiss = 1)	0.06	0.14	0.02	0.05	0.20	0.02	(0.02)
	(0.25)	(0.35)	(0.12)	(0.21)	(0.40)	(0.13)	(0.14)
Pr(only ProtMiss = 1)	0.28	0.19	0.33	0.33	0.24	0.32	0.24
	(0.45)	(0.39)	(0.47)	(0.47)	(0.42)	(0.46)	(0.42)
Pr(ProtCathMiss = 1)	0.14	0.21	0.31	0.04	0.09	0.01	0.23
	(0.34)	(0.41)	(0.46)	(0.20)	(0.29)	(0.09)	(0.42)
# CathMiss	2.90	2.48	4.77	1.57	1.91	2.15	3.12
	(1.65)	(1.30)	(1.69)	(0.69)	(1.38)	(1.60)	(0.95)
# ProtMiss	1.86	1.56	2.88	1.66	1.71	1.43	1.72
	(1.20)	(0.79)	(1.68)	(1.01)	(1.24)	(0.88)	(0.77)
Individual Data							
Female	0.54	0.50	0.54	0.53	0.59	0.52	0.57
	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)
Age	29.12	29.88	29.04	28.61	32.02	29.17	27.77
	(10.11)	(10.79)	(10.20)	(10.12)	(8.14)	(10.37)	(9.93)
Married	0.59	0.63	0.56	0.63	0.60	0.58	0.52
	(0.49)	(0.48)	(0.49)	(0.48)	(0.48)	(0.49)	(0.49)
Employed	0.71	0.69	0.70	0.80	0.84	0.67	0.54
	(0.45)	(0.45)	(0.45)	(0.39)	(0.35)	(0.46)	(0.49)
Electricity	0.21	0.22	0.23	0.09	0.11	0.24	0.35
	(0.40)	(0.41)	(0.42)	(0.28)	(0.31)	(0.42)	(0.47)
Radio	0.60	0.52	0.75	0.59	0.62	0.66	0.52
	(0.48)	(0.49)	(0.42)	(0.49)	(0.48)	(0.47)	(0.49)
TV	0.22	0.19	0.31	0.11	0.06	0.32	0.33
	(0.41)	(0.39)	(0.46)	(0.32)	(0.24)	(0.46)	(0.47)
Bicycle	0.39	0.25	0.34	0.51	0.45	0.47	0.30
	(0.48)	(0.43)	(0.47)	(0.49)	(0.49)	(0.49)	(0.45)
Solid Roof	0.35	0.45	0.10	0.35	0.10	0.49	0.60
	(0.47)	(0.49)	(0.27)	(0.47)	(0.39)	(0.50)	(0.48)
Solid Floor	0.39	0.27	0.45	0.23	0.27	0.44	0.66
	(0.48)	(0.44)	(0.49)	(0.42)	(0.44)	(0.49)	(0.47)
HIV Knowledge	3.95	3.08	4.40	3.91	4.15	4.02	4.09
	(2.12)	(2.32)	(1.93)	(2.18)	(1.91)	(2.05)	(2.09)
Tested HIV	0.33	0.12	0.51	0.64	0.13	0.32	0.22
	(0.47)	(0.32)	(0.49)	(0.47)	(0.34)	(0.46)	(0.41)
Village Data							
Rural	0.69	0.51	0.71	0.86	0.75	0.56	0.67
	(0.46)	(0.49)	(0.45)	(0.34)	(0.43)	(0.49)	(0.46)
Village Size	1.59	2.15	1.59	1.21	1.44	1.66	1.68
	(1.02)	(1.34)	(1.06)	(0.62)	(0.88)	(0.93)	(1.09)
Expl. Route	0.25	0.45	0.40	0.05	0.38	0.42	0.09
	(0.44)	(0.49)	(0.49)	(0.12)	(0.48)	(0.49)	(0.28)
Water	0.48	0.54	0.35	0.44	0.41	0.50	0.49
	(0.49)	(0.47)	(0.47)	(0.49)	(0.49)	(0.50)	0.50
Altitude	1,018	722	1,325	848	1,008	1,152	1,111
	(525)	(469)	(680)	(353)	(525)	(185)	(303)
Railway	0.28	0.28	0.33	0.19	0.21	0.34	0.35
	(0.45)	(0.44)	(0.47)	(0.39)	(0.41)	(0.47)	(0.47)
Road	0.27	0.10	0.45	0.18	0.17	0.38	0.38
	(0.44)	(0.18)	(0.49)	(0.38)	(0.37)	(0.48)	(0.48)
Obs.	42,193	6,016	4,679	9,771	5,818	7,816	8,093

Note: the table reports the variables' mean and their standard deviation (in parentheses).

of the early missionary stations: favorable climatic conditions, access to a source of fresh water and the proximity to an explorer route.

In the empirical analysis I take those determinants into account. If they had a long-term impact on the outcomes of interest, not controlling for them could potentially introduce a bias in the estimated effect of Christian missions.

I use the village's elevation (in meter) as a proxy of its climatic conditions since, on average, a more temperate climate (and the frequency of rainfalls) is positively related to altitude. In order to account for the access to a source of fresh water, I construct a binary variable indicating the presence of inland water (lake or river) within 10 kilometers from the village (this information is taken from the *Digital Chart of the World*). The proximity to an explorer's route wants to capture a form of path dependence. The routes of the first missionary explorers (as David Livingstone, for example) determined which parts of Africa were the most well known to Europeans as well as the location of the earliest mission stations. From these first stations, missionaries could penetrate deeper into the interior while maintaining access to supplies from Europe. Information about the first explorers' routes comes from a map produced by the *Century Company* (1897) which provides an overview of the exploring activity in the African continent as of 1895. The map, used also in Nunn (2010) and Nunn and Wantchekon (2011), is reproduced in Figure 2, with the explorers' routes highlighted in pink. Combining this map with the villages' geographical information I create a binary variable equal to one if a pre-1900 explorer route in a radius of 10 kilometers from the village exists. The bottom panel of Table 3 highlights how about one quarter of the sampled individuals live in a village close to a historical explorers' route (with a larger share in Congo DR, Kenya, Tanzania and Zambia, the countries that have been touched more intensively by the exploration activity during the ninetieth century).

In order to take into account the current level of development, I also control for

the presence of a railway and a primary road in a radius of 10 kilometers from the village: in this case data are from the *Digital Chart of the World*. The DHS do not contain information about villages' population. Nevertheless, it is possible to know if an individual lives in a large city (population larger that one million), in a city (population larger that 50,000) or in a town (population smaller than 50,000). I use these categories in order to control for the size of the location of residence.

Table 4 looks at the predictive power of the above variables in explaining the location of Christian missionaries. As suspected the proximity to an explorer route is positively associated with the village's number of historical Christian mission. The village's elevation (which proxies the presence of a more favorable climate) also has a positive effect on the location of missions (especially Protestant), as well as the vicinity to a source of fresh water (but in this case the effect is not statistically significant).

2.4 The Long-Term Effect of Christian Missions

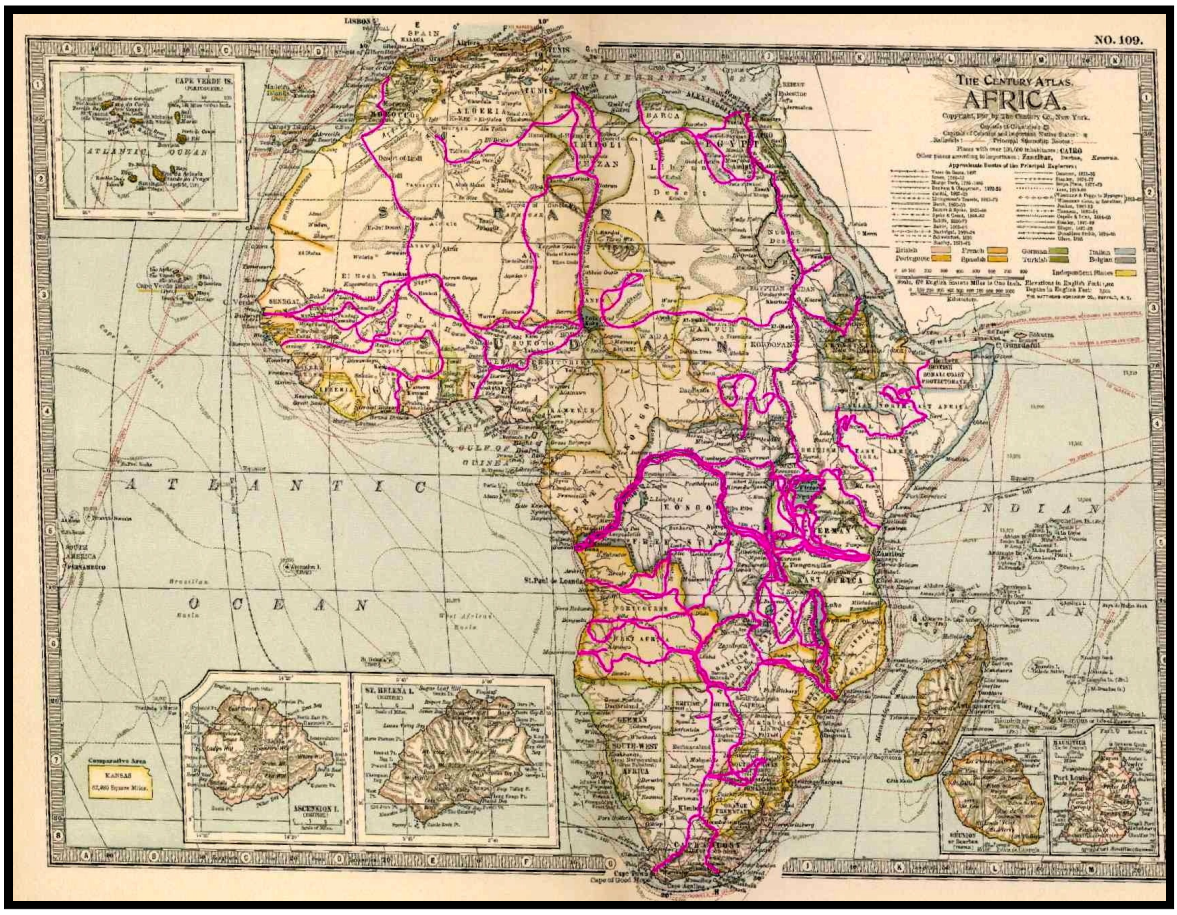
To capture the long-term effect of Christian missions, I estimate the following regression equation

$$\Pr\left(Y_{i,j}=1\right)=\alpha ChristianMissions_j+\mathbf{X_{i,j}}'\beta+\mathbf{Z_j}'\gamma+\sum_{c=1}^{C}\psi_c d_c+\epsilon_{i,j} \qquad (1)$$

where i, j denote individuals and villages respectively; $Y_{i,j}$ indicates the outcome variable; $Christian\ Missions_j$ is equal to the number of Christian missions within a radius of 30 kilometers from the village j at the beginning of the twentieth century; $\mathbf{X_{i,j}}$ is the vector of individual characteristics and $\mathbf{Z_j}$ is the vector of geographical controls at village level; d_c are country dummies controlling for country-specific characteristics; finally $\epsilon_{i,j}$ is an *i.i.d.* error term.

The data allow me to distinguish between Protestant and Catholic missions to

Figure 2: Explorers' Routes in Africa during the 19th Century



Source: *The Century Company* (1897)

test whether there is evidence of any differential effect between the two religious denominations. In this case the baseline equation is

$$\Pr\left(Y_{i,j}=1\right)=\alpha_1 ProtMissions_j+\alpha_2 CathMissions_j+\mathbf{X_{i,j}}'\beta+\mathbf{Z_j}'\gamma+\sum_{c=1}^{C}\psi_c d_c+\epsilon_{i,j} \tag{2}$$

where, now, $ProtMissions_j$ and $CathMissions_j$ indicate, respectively, the number of Catholic and Protestant missionary stations within a radius of 30 kilometers from the the village j as of 1911.

The assumption is that the number of missions during the colonial period is not correlated with the error term $\epsilon_{i,j}$. Since the Christian missions considered in this analysis were established over a century ago, it is reasonable to assume that they are predetermined with respect to individual characteristics that affect the respondent's sexual behaviors.

One possible concern may be related to the existence of omitted variables that both determined the location of the early Christian missionaries during the colonial period and that, having persisted over the decades, also have an effect on current individual sexual outcomes. As explained in the previous section, historians identify specific factors that determined the location of missionary stations in the African continent. My argument is that, controlling for those factors and for both country and regional fixed effects, we can safely assume the exogeneity of Christian (Catholic and Protestant) missions in Equation 2 and 3.

2.4.1 Christian Missions and HIV

Table 5 reports the estimated coefficients from the regression of Equation 2 and Equation 3 when the dependent variable is an indicator equal to one if the individual is HIV positive and zero otherwise.

The results show that, on average, the historical presence of Christian missions

does not have a significant impact on the individual HIV infection rate.

When we distinguish between the two religious denominations, however, an interesting pattern emerges. While Protestant missions do not significantly affect the dependent variable, Catholic missions do. Individuals living in a village that had a higher number of Catholic missionary stations during the colonial period have a lower probability of being HIV positive. The estimated coefficients, along the different specifications, are consistently negative and statistically significant. In particular, once we control for individual and village characteristics and we include both country and regional fixed effects, the results suggest that a one unit increase in the number of historic Catholic missions at village level, decreases the HIV infection rate of current inhabitants by 1.2 percentage points.

To test the heterogeneity of these results, I re-estimate the model separately for men and women, for rural and urban areas and for different levels of education. Table 6 show that the negative impact of historic Catholic mission is stronger for male respondents and for individuals living in urban areas and with a higher level of education.

HIV can be transmitted in a number of ways, including through unprotected sexual intercourse; from mother to child during pregnancy, childbirth or delivery; and through intravenous drug use, an infected blood supply, or contaminated medical equipment. In sub-Saharan Africa, however, the vast majority of HIV infections are transmitted via heterosexual intercourse (UNAIDS [2010]). Since sex represents the main vector for the transmission of the virus, the previous finding may be indicative of a different long-term impact of the two religious groups on the individual probability of adopting safe forms of sexual behavior. This will be the focus of the following analysis.

2.4.2 Christian Missions and Condom Use

The "C" component of the ABC strategy advocates the correct and consistent use of condoms during sexual intercourse as one of the pillars for a safe sexual behavior.

Table 7 reports the estimated long-term effect of Christian missions on the probability of having used a condom during the last sexual intercourse. Overall, the degree of historic exposure to Christianity has a small negative (though not significant) effect on the probability of adopting such protective behavior. Moreover, once we include country and regional fixed effects, the estimate coefficient loses most of its statistical significance.

A very specific patterns, however, emerges when we differentiate between Catholic and Protestant missions. The results in the bottom panel of Table 7 reveal the existence of a differential impact of the two religious denominations. Across the different specifications, living in a village that has been historically exposed to the presence of Catholic missions determines a significantly lower propensity for condom use, while Protestant missions do not result to have any significant impact on that same probability. The baseline specification (column V) suggests that an additional Catholic mission in the village decreases today's inhabitants chance of having used a condom during the last sexual intercourse by 0.7 percentage points. This result, however, is only mildly statistically significant.

One plausible explanation for this finding may be related to the different official positions toward the use of condoms (and, in general, other forms of contraception) held by the Catholic and the Protestant churches. It is well known that the Catholic Church strongly refuses the use of any form of birth control, including condoms, because that would undermine the only purpose of the sexual act: procreation. This dogmatic ban extends also to cases where such use may be instrumental in reducing the spread of sexually transmitted diseases such as HIV. Still in 2009, Pope Benedict XVI stated that the AIDS pandemic is a tragedy that "cannot be overcome through

the distribution of condoms, which even aggravates the problem."[61] More heterogenous positions regarding the use of contraceptions can be found within the different Protestant denominations. Many Protestant groups have instead a more open stance (or, at least, a less strong opposition) toward the use of protections (such as condoms) during sex. The Anglican Church of England, for example, does not regard contraception as a sin or a contravention of God's purpose. The Lambeth Conference of 1958 passed a resolution stating that "the responsibility for deciding upon the number and frequency of children is laid by God upon the consciences of parents in such ways as are acceptable to husband and wife." Ten years later, the Bishops of the Anglican Church officially disagreed with the Papal Encyclical *Humanae Vitae* and its view that methods of contraception are contrary to the will of God. A similar openness toward other forms of contraception is found in the other two traditional Protestant denominations (Calvinists and Lutherans).[62] In the case of Africa, lastly, there are several cases of Protestant bishops and ministries openly speaking in favor of the use of condoms in the fight against HIV.[63]

The effect of Catholic and Protestant missions varies across several sub-samples of the population. The heterogeneity analysis shown in Table 13 reveals how the impact of Catholic missions is different depending on respondents' gender: the negative effect of Catholic mission appear to be stronger for men (with a marginal effect of 1.5 percentage points) relative to women.

[61] *Daily Telegraph*, March 17th 2009. The Catholic Church has traditionally opposed every form of contraception (not limited to condoms only) as a violation of the procreative and unitive meaning of the conjugal relation. The *Catechism of the Catholic Church* (quoting Pope Paul VI's 1968 encyclical *Humanae Vitae*) states that "every action which, whether in anticipation of the conjugal act, or in its accomplishment, or in the development of its natural consequences, proposes, whether as an end or as a means, to render procreation impossible is intrinsically evil" (2370).

[62] For example the Lutheran Church allows for contraception in the event the potential parents do not intend to care for a child

[63] In the words of the Anglican Bishop Karuna of Uganda: "If you are foolish enough to have sex outside [the marriage], don't be stupid enough not to use a condom (reported in Hardee [2008]). Similarly, the Anglican Bishop Solomon Tilewa of Gambia: "We are aware that we live in a world where not everybody is holy and for some people abstinence or one partner is not a viable proposition, therefore, the only sensible and responsible line of action is the use of condoms" (cited from Colombant [2005]).

One possible concern with the findings shown in Table 7 is that the propensity to use a condom depends also on the identity of the sexual partner and the two choices (i.e. identity of the partner and whether or not to use a condom) are likely to be correlated with each other. So, for example, if one decides to have sex with an occasional partner, she might also have a higher propensity to use a condom since the chances of sexual diseases (including AIDS) are higher relative to the case of sexual intercourse with her spouse. At the same time, for two people in a stable relationship, other reasons (such as the desire of having a child) arise to justify the lack of condom use. Let's denote as $P^*_{i,j}$ and $Y^*_{i,j}$ the individual indirect utility function associated, respectively, with choosing her spouse as last sexual partner and the choice of using a condom during the last sexual intercourse. Although I do not observe the individuals' indirect utilities, I do observe their choices so that I can model the problem as a system of dichotomous variables:

$$P_{i,j} = \begin{cases} 1, & \text{if } P^*_{i,j} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$Y_{i,j} = \begin{cases} 1, & \text{if } P^*_{i,j} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

with $P^*_{i,j} = \mathbf{X^{P\prime}_{i,j}}\beta^P + \theta^P_{i,j}$ and $Y^*_{i,j} = \mathbf{X^{Y\prime}_{i,j}}\beta^Y + \theta^Y_{i,j}$.

Under the assumption of $\theta^P_{i,j}$ and $\theta^Y_{i,j}$ being jointly normally distributed with zero mean, unitary variance and a correlation coefficient equal to ρ, I can simultaneously estimate Equation 4 and Equation 5 via a Bivariate Probit. The estimated coefficients in Table 8 confirm the previous results: the Catholic heritage decreases the likelihood of using a condom, even when taking into account the simultaneity between this choice and the identity of the last sexual partner. The correlation coefficient between the error terms of the two equations is negative and largely statistically significant,

meaning that unobservables affecting positively the decision of having sex with the spouse have a negative impact on the decision of whether or not to use a condom.

2.4.3 Christian Missions and Pre-Marriage Sexual Abstinence

The Christian teachings praise sexual abstinence before marriage as a way to become closer to God.[64] Accordingly, Christian hierarchies and faith-based organizations promote sexual abstinence as the most effective way to reduce the chances of being infected by HIV and, therefore, they educate individuals not to have sexual intercourse before marriage. Given the importance attributed to pre-marital abstinence by the Christian doctrine, I expect to find a positive relationship between the dependent variable (probability of having abstained) and the degree of historic exposure to Christianity.

Table 9 does't reveal a significant relationship between historical exposure to Christianity and premarital sexual abstinence. Contrary to the previous case, we don't see any significant differences between Catholic and Protestant missions either. The exposure to Catholic institutions translates, as expected, into a higher individual probability to have abstained from sex before marriage but this relationship is not significant one one takes into account country and regional fixed effects.

In Table 14 I verify the presence of heterogenous effects across different subsamples of the population. The findings reveal that an increase in the village's number of historical Catholic missions determines a significant increase in the likelihood of premarital sexual abstinence for better educated individuals and for men a group that, as evidenced in Table 2, tends to be more sexually active during the pre-marital period.

[64] *Corinthians* (7:2) clearly includes sex before marriage in the definition of sexual immorality ("Now concerning the matters about which you wrote: "It is good for a man not to have sexual relations with a woman" But because of the temptation to sexual immorality, each man should have his own wife and each woman her own husband."). There are numerous other scriptures that declare sex before marriage to be a sin: *Exodus* (22:16), *Acts* (15:20), *Hebrews* (13:4), *Colossians* (3:5), *Thessalonians* (4:3).

Related to this analysis, in Table 10 I consider, as dependent variable, the age of sexual debut. Results show that, conditioning on the age of the first marriage, Protestant missions tend to reduce the age of sexual debut (but this effect is not statistically significant). On the other hand Catholic missions are associated with a delay in the occurrence of the individual first sexual intercourse. To the extent that sexual activity at a young age is positively correlated with sexual promiscuity and, consequently, with a higher infection risk, the results point to the fact that Catholic missions have been more effective, relative to their Protestant counterparts, in promoting preventive forms of sexual behavior.

2.4.4 Christian Missions and Extra-Marital Sex

Christian teachings consider adultery immoral.[65] Given the strong condemnation of marital unfaithfulness by Christianity, I expect a negative impact of Christian missions on this type of sexual behavior. Intuitively the historic presence of Christian institutions may have determined the creation and development of a set of moral values that limit the individual propensity of engaging in this type of behavior.

Table 11 reports the estimated coefficients for Equation 2 and Equation 3 where the dependent variable is a dummy equal to one if the respondents declares to have had sexual intercourse with someone else other than his/her partner during the last twelve months. Existing evidence has highlighted how polygyny may affect the propensity of engaging in extra-marital sex. To take that into account, I include and additional

[65]Several passages in the Old and New Testament makes the point clear. *Exodus* (20:14): "You shall not commit adultery." *Corinthians* (6: 9-10): "Do you not know that the unrighteous will not inherit the kingdom of God? Do not be deceived: neither the immoral, nor idolaters, nor adulterers, nor sexual perverts, nor those who practice homosexuality, nor thieves, nor the greedy, nor revilers, nor robbers will inherit the kingdom of God." *Mark* (10:12): "And if she divorces her husband and marries another man, she commits adultery." *Luke* (16:18): "Anyone who divorces his wife and marries another woman commits adultery, and the man who marries a divorced woman commits adultery." The Catechism of the Catholic Church states that "Adultery refers to marital infidelity [...]. Christ condemns even adultery of mere desire. The sixth commandment and the New Testament forbid adultery absolutely. The prophets denounce the gravity of adultery; they see it as an image of the sin of idolatry."

regressor to indicate if the respondent declares to be part of a polygynous relationship.

I find evidence of a negative relationship between historic Christian missions and marital infidelity: an additional mission in the village determines a 0.2 percentage points decrease of today's individuals' propensity of engaging in marital unfaithfulness. When I distinguish between the two religious denominations, I find that most of this effect is accounted for by Catholic missions. A one unit increase in the village's number of historic Catholic stations is responsible for a 0.8 percentage points decrease of the dependent variable. When I control for both country and regional fixed effects, instead, Protestant missions appear not to have a significant effect (and the corresponding coefficient is also quantitatively smaller relative to Catholic missions). These findings support the initial hypothesis of a negative relationship between marital infidelity and historic Christian heritage and highlight how Catholic missions, in particular, were more effective in creating and developing a system of moral norms that, having persisted over time, limit a potentially risky form of sexual behavior.

It is worth noticing, finally, how the impact of Catholic missions is stronger for those groups that are more at risk of engaging in extra-marital sex (see Table 16). As seen in Table 2 there is a striking gender difference in the likelihood of extra-marital relationships. This is not a specific feature of this sample: empirical evidence has revealed numerous male and female differences in extra-marital sexual attitudes and behaviors triggered, in most cases, by specific gender role expectations.[66] It is therefore significant to see that Catholic missions decrease the probability of engaging in marital unfaithfulness particularly for the males sample, that is for those individuals that are more "at risk" of acquiring HIV outside of marriage.

Individuals living in urban areas may be more able to build extra-marital ties with the people they meet because they handle less social pressure with respect to social

[66]See, for example, Thompson (1983), Schensul et al. (2006) and Smith (2007).

and familial norms than someone living in a village, and also because urbanization makes people more anonymous, which confers them the possibility to do things in secret, such as having multiple patterns. Empirical evidence show how, precisely, extra-marital sex is more likely to happen in an urban environment relative to rural areas. As Table 16 shows, the historical exposure to Catholic missions tends to significantly reduce the propensity of engaging in extra-marital sex among individuals living in an urban setting, representing, therefore, an effective constraint to a highly risky sexual behavior.

2.5 Education, Religiosity or Culture?

The previous findings support the hypothesis that Christian missions established during the colonial period have an effect on AIDS-related sexual behaviors that persisted over decades. This effect can be the result of different factors: education, religiosity and culture.

Empirical evidence has shown how religious institutions can be instrumental in promoting investments in human capital. The establishment of schools (and other educational institutions) as well as the promotion of literacy was (and still is) one of the main functions of the Christian missionary activity. Becker and Woessmann (2009), for example, find that Protestantism was the main driving force behind the high literacy rates and the decreasing female education gap in late ninetieth century Prussia. Similarly, using regional data for 180 African provinces, Gallego and Woodberry (2010) find a positive relationship between historical Protestant missionary activity and current schooling variables.

Equivalent results are found in this analysis, too. Columns I and II of Table 17 in the appendix report the estimates from a regression where the dependent variable is represented by the individual years of completed education. The first column suggests that individuals residing in a village that has been historically exposed to the presence

of Christian missions display a higher level of education.[67] Confirming the findings in Gallego and Woodberry (2010), I also find evidence of a different role played by Protestant and Catholic missionaries in promoting human capital. One reason for this result can be related to theological differences between the two religious denominations. Protestantism (Lutheranism in particular) advocates the individual ability of comprehending God's Word without the intermediation of priests or presbyteries (as opposed to the Catholic practice). This view is embodied in the concept of *Sola Scriptura*, considered by some Protestant groups to be one of the theological pillars of the Reformation. The key implication of the principle is that interpretations and applications of the Scriptures do not have the same authority as the Scriptures themselves; hence, the ecclesiastical authority is viewed as subject to correction by the Scriptures, even by an individual member of the Church. With the words of Martin Luther: "a simple layman armed with Scripture is greater than the mightiest pope without it." The intention of the Reformation was to correct the perceived errors of the Catholic Church by appealing to the uniqueness of the Bible's authority and to reject what Catholics considered to be Apostolic Tradition as a source of original authority alongside the Bible, wherever Tradition did not have biblical support or where it supposedly contradicted Scripture. Quite obviously, if one wants to read the Bible, one must be able to read it; Protestant missionaries have therefore the explicit duty of promoting literacy in order to give everyone the ability to read and understand the Holy Scriptures. This represented the main incentive for Protestant missionaries to found schools.

Education, in turn, can be correlated with individual sexual behavior. It is likely, for example, that more educated individuals have better access to information regarding protective sexual practices (as condom use). Analogously, it may be that

[67] I also performed the same analysis using, as dependent variable, a dummy defining low (i.e. less than primary) and high (i.e. primary and beyond) education. The results are qualitatively equivalent.

individuals who are in school tend to postpone marriage and result, therefore, to be more likely to experience pre-marital sexual relationships. In order to isolate the effect of Christian missions on individual sexual behavior via education, in column I of Table 12 I repeat the previous analysis including the years of completed schooling as additional explanatory variable. Notice, first, how education has, as expected, a positive effect on the probability of condom use and a negative effect on the probability of having abstained from sex before marriage; more educated individuals, moreover, appear to be less likely to engage in extra-marital sexual relationships. For all the three types of sexual behavior, the inclusion of this additional variable only slightly decreases the magnitude (in absolute value) of the coefficient attached to Protestant and Catholic missions without affecting their statistical significance in a substantial way. Even when controlling for the, possibly confounding, effect of education I find confirmation of the long-term effect of Christian missions on today's individuals sexual behavior.

Missions may also affect sexual behaviors through their effect on individual religiosity. Column IV of Table 17 in the appendix shows that an increase in the village's number of Protestant and/or Catholic missions is associated with a higher probability of today's residents declaring to belong to a Christian denomination. Religion is thought to provide people with sets of moral teachings about what is right and wrong and how they, therefore, ought to live. Both Catholic and Protestant teachings advocate the refusal of certain types of behavior (condom use, marital unfaithfulness) and the praise of certain others (pre-marital sexual abstinence). Individuals belonging to those religious denominations will be more apt to reflect in their actions the implications of those moral teachings. Column II of Table 12 aims to see how much of the long-term effect of Protestant and Catholic missions has to be attributed to this religiosity "channel". The answer seems to be not that much. Even when controlling for the religious denomination individuals belong to (Christians, Muslims, Others/None),

the coefficients attached to Protestant and Catholic historical missions are not heavily affected in either their magnitude or statistical significance.[68]

It seems, therefore, that historical missions affect sexual behaviors through channels that go beyond education and individual religiosity. This is not surprising as individual behaviors (and, sexual practices in particular) are often influenced by the moral and ethic views of the community the single individual belongs to. Peer effects, in particular, may play an important role. Conditioning on the same level of education and the same religious belief, an individual living in a relatively more liberal society may adopt behaviors that are very different from those adopted by someone living in a more conservative community. It is likely that villages that have seen the presence of Christian missions for several decades have developed a set of moral values that persisted over time representing the main source of a peer effect which influences sexual behaviors independently of individual religious beliefs. So, for example, a girl living in an "historically Catholic" community may be less likely to engage in pre-marital sex not just because of her religiosity but, rather, because of the fear of a negative judgement ("social stigma") by her peers.

Finding a measure of peer effect is not straightforward. In this paper I use the village's share of individuals adopting the particular sexual behavior under consideration. When considering condom use at last sexual intercourse as dependent variable, for example, my proxy of peer effect is given by the village's (weighted) share of people who declare to have used a condom.[69] Column III of Table 12 adds this variable among the controls. As we see, accounting for peer effects significantly reduces the impact of Protestant and Catholic missions on all the three types of sexual behaviors,

[68]It would be interesting to see whether the intensity of the religious beliefs has an effect on sexual behaviors, too. It is likely, indeed, that an observant Catholic displays behaviors that are different from those of a non-observant Catholic. In the literature, religious intensity is usually measured with the frequency of participation to Mass (or other equivalent religious liturgies). Unfortunately the DHS do not provide this kind of information.

[69]I am aware that this is far from being a perfect measure of the peer effect. However it represents the best available proxy.

making the corresponding estimated coefficients statistically insignificant. I interpret this result as suggestive of the fact that most of the long-term impact of historical missions on individual sexual behavior is due to the development and persistence across generations of a set of moral norms that shaped the cultural environment at village/community level leading its residents to adopt certain sexual behaviors as the result of social influence and peer effects.

2.6 Conclusions

Though its impact on values and culture, religion can persistently affects individual choices and behaviors. in this paper I explain current differences in individual HIV-related sexual behaviors in Sub-Saharan Africa by exploiting, as a quasi-natural experiment, variations in the exposure to Christianity at the beginning of the twentieth century. Historical exposure to Christianity is measured by using an original dataset containing information about the location of Christian missions at the beginning of the twentieth century. The results of the paper show a long-term negative effect of Catholic missions on the HIV infection rate in the region. Protestant missions, on the other hand, do not appear to have any significant impact. Given that sex represents the main vector of HIV transmission in the area, this finding may be indicative of the fact that the two religious denominations have different effects on individual sexual behaviors. The results show how the Catholic heritage leads people to a lower propensity to having used a condom during the last sexual intercourse but, at the same time, to a higher likelihood to adopt safer forms of preventive behaviors (pre-marriage sexual abstinence and sexual faithfulness). This is in line with recent medical research (see for example Hearst and Chen [2004]; Shelton [2007]; Potts *et al.* [2008]) showing that, in the case of sub-Saharan Africa, there is no evidence of a primary role of condom use in generating a measurable slowing of new infections and that prevention of promiscuous behaviors (such as partner reduction) have resulted

to have greater effect in the fight against AIDS.

2.7 Tables

Variables' Definition

Outcome Variables	
HIV^+	dummy = 1 if HIV positive. Source: DHS.
Condom Use	dummy = 1 if a condom was used during last sexual intercourse.
Abstinence	dummy = 1 if individual has waited until marriage to have sex for the first time. Source: DHS.
Age at First Sex	age (in years) at first sexual intercourse. Source: DHS.
Extra-Marital	dummy = 1 if individual had extra-marital sex in the last 12 months. Source: DHS.
Individual Controls	
Age	age in years. Source: DHS.
Female	dummy = 1 if female. Source: DHS.
Married	dummy = 1 if married or in a stable relationship and living with partner.
Employed	dummy =1 if employed in last 12 months. Source: DHS.
Education	years of education. Source: DHS.
Tested	dummy = 1 if ever tested for HIV before. Source: DHS.
Know HIV	dummy = 1 if know someone with AIDS or died because of AIDS. Source: DHS.
HIV Knowledge	index measuring the degree of knowledge about HIV/AIDS. It ranges from - 6 (lowest knowledge) to + 6 (highest knowledge). Source: DHS.
Electricity	dummy = 1 if household has electricity in the house. Source: DHS.
Radio	dummy = 1 if household owns a radio. Source: DHS.
TV	dummy = 1 if household owns a TV. Source: DHS.
Bicycle	dummy = 1 if household owns a bicycle. Source: DHS.
Refrigerator	dummy = 1 if household owns a refrigerator. Source: DHS.
Car	dummy = 1 if household owns a car. Source: DHS.
Roof Hard	dummy = 1 if household lives in a house with roof made with hard material. Source: DHS.
Floor Hard	dummy = 1 if household lives in a house with floor made with hard material. Source: DHS.
Village Controls	
Urban	dummy = 1 if in urban area. Source: DHS.
Village Size	Categories: large city (pop. $>$ 1,000,000); city (50,000 $<$ pop. $<$ 1,000,000); town (pop. $<$ 50,000). Source: DHS.
Explorer Route 1895	dummy = 1 if a pre-1895 explorer route was in a 10 km radius from the cluster. Source: *The Century Company* (1897); own calculations.
Altitude	cluster's altitude in meters. Source: DHS.
Water	dummy = 1 if a source of fresh water (lakes, river, water basin) exists in a 10 km radius from the cluster. Source: *The Digital chart of the World*; own calculations.
Road	dummy = 1 if a primary road exists in a 10 km radius from the cluster. Source: *The Digital chart of the World*; own calculations.
Railroad	dummy = 1 if a primary railroad exists in a 10 km radius from the cluster. Source: *The Digital chart of the World*; own calculations.

Table 4: Where did the Missionaries Locate?

	Christian Missions	**Catholic Missions**	**Protestant Missions**
Explorer Route	0.62***	0.26***	0.36***
	(0.09)	(0.03)	(0.07)
Road	1.22***	0.27***	0.95***
	(0.09)	(0.03)	(0.08)
Railway	0.57***	0.26***	0.30***
	(0.08)	(0.03)	(0.07)
Lake	-0.029	0.07***	-0.101**
	(0.06)	(0.03)	(0.05)
River	-0.15*	-0.081***	0.069
	(0.06)	(0.02)	(0.05)
Altitude	0.0003***	-0.0005	0.003***
	(0.00007)	(0.0003)	(0.00006)
Country Dummies	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes
N	2,557	2,557	2,557
R-squared	0.25	0.61	0.20

Note: Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1. 5 ad 10 percent.

Table 5: Christian Missions and HIV Infection

Dep. Var.: Pr(HIV positive)$_{i,j}$ = 1

	I	II	III	IV	V
# Christian Miss.	0.013	-0.027**	-0.040***	-0.007	0.006
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
# Catholic Miss.	-0.040	-0.132***	-0.113***	-0.103***	-0.068**
	(0.027)	(0.028)	(0.029)	(0.030)	(0.031)
# Protestant Miss.	0.032**	0.007	-0.021	0.016	0.023
	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)
Observations	42,193	42,193	42,193	42,193	42,193
Individual Controls	No	Yes	Yes	Yes	Yes
Village Controls	No	No	Yes	Yes	Yes
Country Dummies	No	No	No	Yes	Yes
Regional Dummies	No	No	No	No	Yes

Note: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge.
"Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.

Table 6: Catholic Missions and HIV Infection: Heterogeneity

	Males	Females	Urban	Rural	Low Ed.	High Ed.
# Catholic Miss.	-0.094**	-0.052	-0.096*	-0.036	-0.051	-0.069*
	(0.041)	(0.035)	(0.051)	(0.041)	(0.047)	(0.038)
# Protestant Miss.	0.011	0.035**	0.030	0.040**	0.028	0.021
	(0.019)	(0.015)	(0.024)	(0.016)	(0.020)	(0.017)
Observations	19,403	22,790	12,328	29,955	18,664	23,529
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the Probit estimated coefficient for Catholic missions only when estimating column V over different sub-samples of the population.
"Low Education" refers to individuals with no formal education or with only incomplete primary education. "High Education" refers to individuals with a level of education beyond completed primary education. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.

Table 7: Christian Missions and Condom Use

Dep. Var.: $\Pr(\text{Condom Use})_{i,j} = 1$

	I	II	III	IV	V
# Christian Miss.	0.045***	-0.019**	-0.026***	-0.005	-0.002
	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)
# Catholic Miss.	0.016	-0.062***	-0.069***	-0.040*	-0.041*
	(0.019)	(0.020)	(0.024)	(0.024)	(0.027)
# Protestant Miss.	0.055***	-0.004	-0.014	0.003	0.007
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	42,193	42,193	42,193	42,193	42,193
Individual Controls	No	Yes	Yes	Yes	Yes
Village Controls	No	No	Yes	Yes	Yes
Country Dummies	No	No	No	Yes	Yes
Regional Dummies	No	No	No	No	Yes

Note: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge.
"Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.

Table 8: Condom Use and Sexual Partner Identity

	Sexual Partner = Spouse	Condom Use = 1
# Catholic Miss.	0.011	-0.058**
	(0.041)	(0.025)
# Protestant Miss.	0.002	0.003
	(0.019)	(0.010)
Obs.	34,450	34,450
Individual Controls	Yes	Yes
Village Controls	Yes	Yes
Country Dummies	Yes	Yes
Regional Dummies	Yes	Yes
ρ		-0.639
		(0.030)
H_0: $\rho = 0$ (p-value)		0.000

Note: estimated coefficients from a Bivariate Probit regression. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls" and "Village Controls" defined as in the notes to Table 7.

Table 9: Christian Missions and Pre-Marriage Sexual Abstinence

Dep. Var.: Pr(Pre-Marriage Sexual Abstinence)$_{i,j} = 1$

	I	II	III	IV	V
# Christian Miss.	-0.006	-0.004	0.003	0.003	0.004
	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
# Catholic Miss.	0.035*	0.037*	0.089***	0.034	0.039
	(0.020)	(0.022)	(0.025)	(0.027)	(0.025)
# Protestant Miss.	-0.020*	-0.017	-0.017	-0.003	-0.003
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Obs.	24,940	24,940	24,940	24,940	24,940
Individual Controls	No	Yes	Yes	Yes	Yes
Village Controls	No	No	Yes	Yes	Yes
Country Dummies	No	No	No	Yes	Yes
Regional Dummies	No	No	No	No	Yes

Note: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge.
"Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.

Table 10: Christian Missions and Age at First Sex

Dep. Var.: Age at First Sex					
	I	II	III	IV	V
# Christian Miss.	0.081***	0.014	0.043*	0.047**	0.046**
	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)
# Catholic Miss.	0.236***	0.073	0.284***	0.125**	0.113**
	(0.051)	(0.050)	(0.052)	(0.050)	(0.052)
# Protestant Miss.	0.030	-0.003	-0.015	0.029	0.032
	(0.024)	(0.025)	(0.026)	(0.025)	(0.025)
Obs.	24,940	24,940	24,940	24,940	24,940
Individual Controls	No	Yes	Yes	Yes	Yes
Village Controls	No	No	Yes	Yes	Yes
Country Dummies	No	No	No	Yes	Yes
Regional Dummies	No	No	No	No	Yes

Note: OLS estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls" and "Village Controls" defined as in the notes to Table 9.

Table 11: Christian Missions and Extra Marital Sex

Dep. Var.: Pr(Extra Marital Sex)$_{i,j} = 1$

	I	II	III	IV	V
# Christian Miss.	-0.034**	-0.031*	-0.040**	-0.042**	-0.038**
	(0.016)	(0.017)	(0.018)	(0.016)	(0.016)
# Catholic Miss.	0.065*	0.070*	-0.015	-0.124***	-0.089**
	(0.039)	(0.042)	(0.047)	(0.044)	(0.044)
# Protestant Miss.	-0.043**	-0.040**	-0.042**	-0.035**	-0.034**
	(0.017)	(0.018)	(0.018)	(0.016)	(0.016)
Obs.	35,466	35,466	35,466	35,466	35,466
Individual Controls	No	Yes	Yes	Yes	Yes
Village Controls	No	No	Yes	Yes	Yes
Country Dummies	No	No	No	Yes	Yes
Regional Dummies	No	No	No	No	Yes

Note: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls": age, age squared, female, employed in last 12 months, married, in a polygynous relationship, household wealth, ever tested for HIV, HIV knowledge.
"Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, urban, village size.
Notice that in this case the sample is restricted to individuals that are currently married.

Table 12: Education, Religiosity and Peer Effect

	Condom Use			Abstinence			Extra Marital Sex		
	I	II	III	I	II	III	I	II	III
# Catholic Miss.	-0.039*	-0.040*	-0.005	0.036	0.036	0.031	-0.093**	-0.080*	-0.057
	(0.017)	(0.017)	(0.018)	(0.025)	(0.025)	(0.024)	(0.045)	(0.044)	(0.034)
# Protestant Miss.	0.006	0.005	0.013	-0.002	-0.001	0.017	-0.029*	-0.028*	-0.021
	(0.012)	(0.012)	(0.010)	(0.012)	(0.012)	(0.011)	(0.016)	(0.017)	(0.016)
Education (Years)	0.018***	0.018***	0.017***	-0.018***	-0.018***	-0.012***	-0.011**	-0.010**	-0.009**
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Religion: Christian		0.026	0.004		0.009	0.030		-0.123**	-0.102**
		(0.043)	(0.041)		(0.032)	(0.034)		(0.050)	(0.049)
Religion: Muslim		0.039	0.019		0.047	0.104*		-0.061	-0.062
		(0.060)	(0.058)		(0.059)	(0.059)		(0.070)	(0.066)
Peer Effect			4.897***			2.796***			3.193**
			(0.219)			(0.110)			(0.126)
Observations	42,193	42,193	42,193	24,940	24,940	24,940	34,642	34,642	34,642
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent. "Individual Controls" and "Village Controls" are defined as in the notes to Table 7, Table 9 and Table 11.

2.8 Appendix A: Additional Tables

Table 13: Missions and Condom Use: Heterogeneity

	Males	Females	Rural	Urban	Low Ed	High Ed
# Catholic Miss.	-0.090***	0.010	-0.029	-0.050	-0.038	-0.032
	(0.034)	(0.037)	(0.036)	(0.042)	(0.051)	(0.031)
# Protestant Miss.	0.0135	0.001	0.014	0.013	0.011	0.006
	(0.015)	(0.015)	(0.014)	(0.021)	(0.016)	(0.014)
Observations	19,403	22,790	29,955	12,238	18,664	23,529
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: the Table reports the estimated coefficients from Probit regressions over different sub-samples of the population. All the specifications include country fixed effects and regional fixed effects. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent. "Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge. "Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.
"Low Education" refers to individuals with no formal education or with only incomplete primary education. "High Education" refers to individuals with a level of education beyond completed primary education.

Table 14: Missions and Pre-Marriage Sexual Abstinence: Heterogeneity

	Males	Females	Rural	Urban	Low Ed	High Ed
# Catholic Miss.	0.074*	0.013	-0.018	0.040	-0.032	0.072**
	(0.038)	(0.032)	(0.033)	(0.047)	(0.041)	(0.031)
# Protestant Miss.	-0.010	-0.002	0.0002	-0.025	0.013	-0.016
	(0.019)	(0.016)	(0.014)	(0.027)	(0.017)	(0.017)
Observations	13,148	11,792	17,601	7,339	10,572	14,368
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: the Table reports the estimated coefficients from Probit regressions over different sub-samples of the population. All the specifications include country fixed effects and regional fixed effects. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent. "Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge. "Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size. "Low Education" refers to individuals with no formal education or with only incomplete primary education. "High Education" refers to individuals with a level of education beyond completed primary education.

Table 15: Missions and Age at First Sex: Heterogeneity

	Males	Females	Rural	Urban	Low Ed	High Ed
# Catholic Miss.	0.136	0.111*	0.099	0.200**	0.0701	0.184***
	(0.082)	(0.062)	(0.075)	(0.099)	(0.083)	(0.065)
# Protestant Miss.	0.040	0.053*	0.055*	-0.015	0.039	0.031
	(0.040)	(0.027)	(0.032)	(0.044)	(0.035)	(0.035)
Observations	13,148	11,792	17,601	7,339	10,572	14,368
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: the Table reports the estimated coefficients from Probit regressions over different sub-samples of the population. All the specifications include country fixed effects and regional fixed effects. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent. "Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge. "Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.
"Low Education" refers to individuals with no formal education or with only incomplete primary education. "High Education" refers to individuals with a level of education beyond completed primary education.

Table 16: Missions and Extra-Marital Sex: Heterogeneity

	Males	Females	Rural	Urban	Low Ed	High Ed
# Catholic Miss.	-0.103*	-0.034	-0.051	-0.201**	-0.083	0.077
	(0.051)	(0.069)	(0.052)	(0.096)	(0.065)	(0.054)
# Protestant Miss.	-0.012	-0.073**	-0.036*	-0.022	-0.008	-0.0391**
	(0.018)	(0.036)	(0.021)	(0.029)	(0.027)	(0.020)
Observations	14,911	19,731	25,636	9,006	16,210	18,432
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: the Table reports the estimated coefficients from Probit regressions over different sub-samples of the population. All the specifications include country fixed effects and regional fixed effects. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent. Individual Controls: age, age squared, female, employed in last 12 months, married, household wealth, ever tested for HIV, HIV knowledge. Village Controls: altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.
"Low Education" refers to individuals with no formal education or with only incomplete primary education. "High Education" refers to individuals with a level of education beyond completed primary education.

Table 17: Effect of Historical Christian Missions on Current Education and Religiosity

	Dep. Var.: Years of Education		Dep. Var.: $\Pr(\text{Christian})_{i,j} = 1$	
	I	II	III	IV
# Christian Miss.	0.059**		0.098***	
	(0.025)		(0.018)	
# Catholic Miss.		-0.150**		0.089**
		(0.062)		(0.036)
# Protestant Miss.		0.107***		0.101***
		(0.028)		(0.021)
Obs.	58,783	58,783	58,783	58,783
Individual Controls	Yes	Yes	Yes	Yes
Village Controls	Yes	Yes	Yes	Yes
Country Dummies	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes

Note: Column I and II: OLS estimates. Column III and IV: Probit estimates. Robust standard errors, clustered at village level, are reported in parentheses. ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
"Individual Controls": age, age squared, female, employed in last 12 months, married, household wealth.
"Village Controls": altitude (in meters), proximity to an explorer route in 1895, proximity to inland water, proximity to a primary road, proximity to a railway, living in rural area, village size.

2.9 Appendix B: Additional Figures

Figure 3: Detail of a map from the *Statistical Atlas of Christian Missions* (places with a Protestant mission are underlined in red)

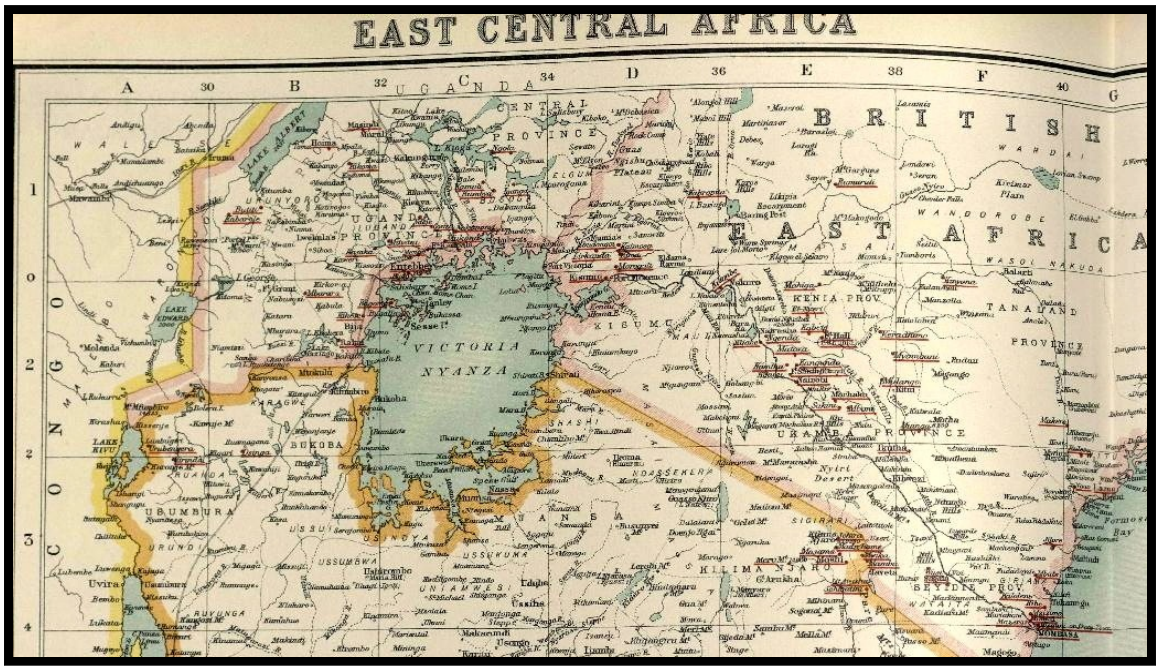
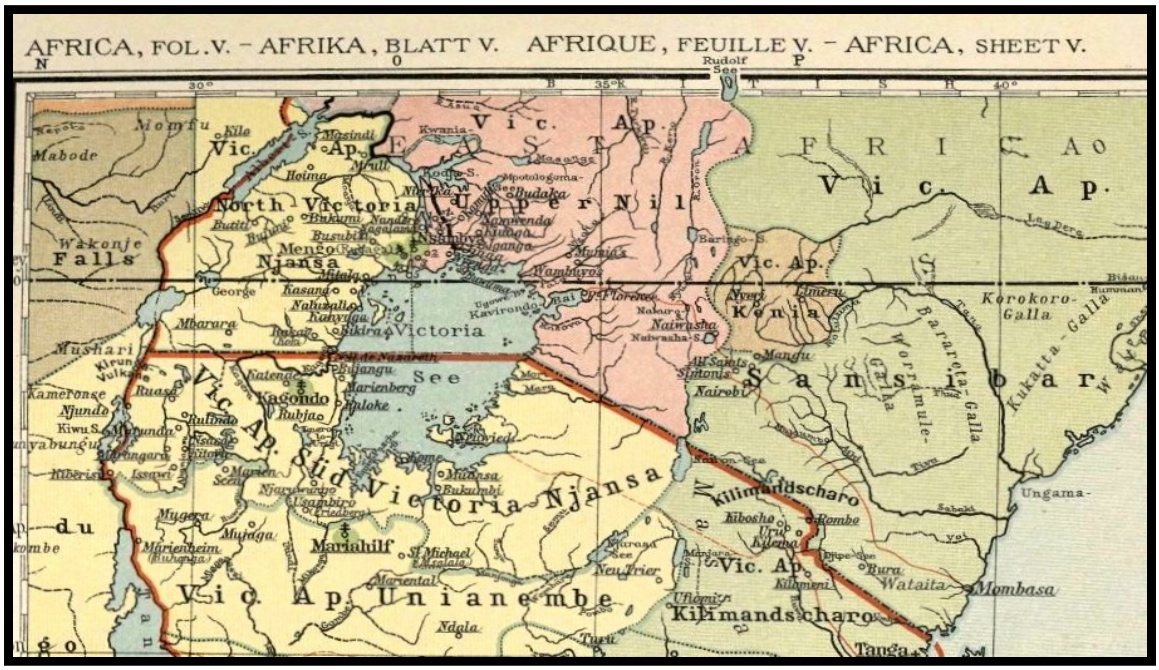


Figure 4: Detail of a map from the *Atlas Hierarchicus* (places with a Catholic mission are underlined)



3 Do Remittances Affect the Labor Supply Decisions of Those Left Behind? Evidence from Albania.

3.1 Introduction

In the last 40 years, the number of international migrants has been constantly increasing. In 2009, 215 million people - about 3 percent of the world population - lived and worked outside their country of birth (UNDP [2009]). Given the global demographic transition and the widening economic gaps between the developed and the developing world, international migration is likely to further increase in the future.

The rise in migration has caused a steady increase of migrants' remittances over the years. In 2010 the global value of remittances amounted to U.S.$ 325 billion and the flows are projected to rise to U.S.$ 374 billion in 2012 (Mohapatra *et al.* [2010]). For most developing countries, migrants' remittances have become the second most important source of external finance after Foreign Direct Investments (FDI). Despite their increasing importance, the impact of international migration and remittances on the economy of the migrants sending countries, is still not settled. At macroeconomic level, findings are mainly inconclusive, ranging from a non-significant to an only mild effect of remittances on economic growth.[70] One possible explanation is that remittances to developing countries are mostly used for consumption, rather than investment in physical and/or human capital, so that they potentially have a short-run impact on growth, rather than a long-run effect. Another reason may be that, since remittances represent direct income transfers, recipients may react by substituting labor with leisure and therefore decreasing their labor supply, either in terms of a lower amount of effort (*intensive margin*), or in terms of a lower participation in

[70] See, for example, Chami et al. (2003), Catrinescu (2009), Giuliano and Ruiz-Arranz (2009), Singh et al. (2010).

the labor market (*extensive margin*).

This paper aims to estimate the magnitude of the labor substitution effect of international remittances using individual and household data from Albania.

At microeconomic level, starting from the mid-nineties, researchers began to analyze the effect of international migration and remittances on several dimensions: poverty and income inequality, education, entrepeneurship, rural development and social capital.[71]

As concerns the analysis of the relationship between international migration, remittances and labor supply, the first studies are Funkhouser (1992) on Nicaragua and Rodriguez and Tiongson (2001) on Philippines. They find a negative effect on male labor force participation while no significant impact is detected for women. These seminal papers, however, completely abstract from the econometric challenges that such estimation poses so that their findings are likely to be biased.

Obtaining a correct estimate of the causal impact of remittances on labor supply is complicated by the presence of endogenous self-selection and reverse causality. Self-selection is due to the fact that a necessary condition in order to receive international remittances is to have a family member living abroad. Households do not randomly select into the migrant/non-migrant status, hence the self-selection problem. Moreover, the migration decision is likely to be determined by some unobservable characteristics that shape also the individual labor supply decisions; hence the endogeneity of the self-selection process. The second problem is represented by reverse causality, which arises whenever a domestic negative labor supply shock triggers a higher demand for

[71] Concerning the relationship between migration/remittances and income distribution see, for example, Adams (1992), Taylor (1992), Taylor and Wyatt (1996), Chiquiar and Hanson (2005), McKenzie and Rapoport (2007). The remittances-education nexus is examined in Hanson and Woodruff (2002), Cox Edward and Ureta (2003), Lopez-Cordoba (2005), McKenzie and Rapoport (2011), Yang (2008). The effect of past migration experience and international remittances with entrepreneurial activity is studied in McCormick and Wahba ([2001], [2003]), Dustmann and Kirchkamp (2002), Mesnard (2004), Woodruff and Zenteno (2007). Finally, Lucas (1987), Rozelle et al. (1999), Perez-Armendariz and Crow (2010) and Pfutze (2012) look at the impact on rural development and social capital.

remittances by the family members left behind, so that the direction of causality is now reversed.

Most research offers a solution to the reverse causality problem by using instrumental variables. Amuedo-Dorantes and Pozo (2006) on Mexico, for example, suggest, as a valid instrument for remittances, the number of Western Union offices in the district where the individuals live, interacted with the household educational composition. Kim (2007), exploiting the longitudinal dimension of his data, uses lagged values of remittances; Justino (2010) instruments remittances with the proportion of recipient households in the community where the individual lives; Acosta (2006) employes village and household networks; Yang (2008) relies on the variation in the exchange rates of the countries where the migrants live.

Even ignoring potential concerns related to the validity of the instruments used,[72] the previous papers do not take into account the problem represented by the self-selection into the pool of *migrant* households (and, therefore, of remittances' recipients) and its endogeneity with respect to the individual labor choices.

One notable exception is represented by Acosta (2006) who acknowledges the existence of the self-selection problem (in addition to the reverse causality one) and suggests the use of Propensity Score Matching (PSM) as a possible solution. The validity of PSM is based, though, on the assumption that selection into migration is driven only by observable individual (and household) characteristics. As pointed out earlier, it is plausible that the selection process is instead driven by unobservable factors that harm the appropriateness of a strategy based on PSM.

The contribution of the present paper to the remittances-labor supply debate is twofold. First, I propose a multi-stage estimation strategy (based on the combination of a selection model and the use of instrumental variables) in order to correct for

[72]For example, the instrumentation strategy in Amuedo-Dorantes and Pozo (2006) is hardly strictly exogenous since it is based on a variable (education) which is likely to determine the individual labor market outcomes, too.

both the endogenous self-selection into the migration status and the reverse causality of remittances. To the best of my knowledge such strategy represents a novelty in the analysis of the remittances-labor supply nexus. I find that, not correcting for the endogenous self-selection into the pool of migrants (and therefore into the pool of potential remittances' receivers) would introduce an upward bias in the estimates, given the positive correlation between unobservables that affect the labor supply choices and the migration decision. Further, I instrument remittances using a set of variables built upon information collected in the foreign countries where the migrants reside. These variables represent *push factors* that are likely to influence the migrants' propensity to remit (and the amount of resources remitted), without being correlated with factors that affect the labor supply decisions of the individuals left behind.

My analysis is based on data from Albania. This country represents a very interesting case study since, after the political and economic collapse occurred in the early 1990s, it experienced a sudden and dramatic emigration process. The analysis of the effects of this *diaspora* on the Albanian economy and living conditions has been at the center of empirical works that looked at the consequences on rural development and agricultural productivity (Carletto *et al.* [2010]; McCarthy *et al.* [2007]), entrepreneurship (Carletto *et al.* [2009]; Piracha and Vadean [2009]) and human capital (Dabalen and Miluka [2008]).

I find that international migration negatively affects the individual labor force participation and the response is stronger for women, for whom having a family member living abroad decreases the participation in the labor market by 23 percent, relative to a 13 percent decrease for men. The amount of annual remittances receipts generates a substitution effect away from the labor market too. In particular I find that a one percent increase in annual remittances receipts determines a decrease in the probability of participating the labor market by 0.01 percent for males and by 0.05

percent for females. The impact at the intensive margin does not appear to be as robust, being negative, as expected, but only mildly significant. Moreover, the negative effect of remittances is particularly concentrated among the older, less educated individuals, residing in urban areas.

The remainder of the paper is organized as follows. Section 2 will illustrate the empirical model at the core of the analysis, the source(s) of potential bias(es) and the adopted estimation strategy. Section 3 will describe the data sources and the construction of the main variables used in the analysis. Section 4 presents and describes the variables used as instruments. Section 5 will discuss the main findings of the paper. Section 6 will, finally, conclude.

3.2 Empirical Model

Obtaining a consistent estimate of the effect of international migration and remittances' receipts on labor supply poses some econometric challenges related to the endogeneity of both regressors. In the following paragraphs I will present the empirical models to be estimated, the problems related to their estimation and the proposed solutions in order to overcome such biases.

3.3 International Migration and Labor Supply

I first start by examining the relationship between labor force participation and international migration.

The choice model is described by the following latent variable specification:

$$L^* = \alpha'_X X_L + \alpha_M M + \epsilon_L \tag{1}$$

where L^* represents the individual net benefit from the participation to the labor market, X_L is a K-vector of exogenous variables and M is the main (binary) variable of interest, taking value equal to one if the respondent has a family member living abroad and zero otherwise.

Agents will decide to enter the labor market only if the expected net benefits are positive. Therefore, the probability that an individual is performing some work activity is:

$$\text{Prob}(L = 1) = \text{Prob}(\alpha'_X X_L + \alpha_M M + \epsilon_L > 0) = \Phi(\alpha'_X X_L + \alpha_M M) \tag{2}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Under the assumption of exogeneity of the variable M, Equation 2 can be estimated via, for example, a Probit model in order to obtain a consistent estimate of α_M, the effect of international migration on the individual labor choice.

It is however likely that the migration variable is correlated with the error term ϵ_L, leading to a biased estimate of α_M. As pointed out by, for example, Lucas and Stark (1985) and Hoddinot (1994), the migration decision is the result of a utility maximization problem solved by both the prospective migrant and the family members left behind. That is, we can model the migration decision M as the outcome of the latent variable model:

$$M^* = \beta'_X X_M + \epsilon_M \tag{3}$$

where X_M is a N-vector of independent variables and ϵ_M is a random error term, so that

$$M = \begin{cases} 1, & \text{if } M^* > 0 \\ 0, & \text{otherwise} \end{cases}$$

A household will send a member abroad only if the net benefits are positive; i.e. if $M^* > 0$.

To allow for the possibility that the unobserved determinants of the individual labor supply decisions and the unobserved determinants of a family's decision to send a member abroad are correlated, I assume the following linear expression for the two error terms:

$$\epsilon_M = \rho_M u + \omega_M$$

$$\epsilon_L = \rho_L u + \omega_L$$

where ω_j is *i.i.d.* with $E(\omega_j u) = 0$ (for $j = M, L$) and $E(\omega_M \omega_L) = 0$. The unobserved term u affects both the the migration decision and the labor outcome, so that

$$E(\epsilon_M \epsilon_L) = \rho_M \rho_L \sigma_u^2 \neq 0$$

where $\sigma_u^2 = E(u^2)$.

Not taking this into account would determine a biased $\hat{\alpha}_M$. The direction of the bias depends on the sign of ρ_M and ρ_L.

Suppose, for example, that u represents the unobserved average level of ability within the household (or the average risk propensity) which affects positively both the propensity to send a family member abroad ($\rho_M > 0$) and the propensity to work ($\rho_L > 0$); in this case $E(\epsilon_M \epsilon_L) > 0$ and we would obtain an upward biased estimate of the effect of international migration on labor supply. But one can find cases where the two error terms are negatively correlated, resulting in a downward bias, so that it is not possible to establish, *a priori*, whether a LS estimate of α_M in Equation 2 would be upward or downward biased.

Potentially, there would be several ways in order to correct for the endogeneity of M in Equation 2. If one discarded the binary nature of the endogenous regressors, one could estimate Equation 2 using IV-Probit or control function techniques. However, as clearly stated in Wooldridge (2002), these estimators require the endogenous variable to be continuous and therefore they can't be applied in this context.

Because both decisions in the model are dichotomous, there are four possible states of the world ($L = [0, 1]$ and $M = [0, 1]$). Under the additional assumption of ϵ_L and ϵ_M being bivariate normal distributed,[73] the likelihood function corresponding to this set of events is a bivariate probit and the model represented by Equation 2 and Equation 3 can be consistently estimated using a *Recursive Bivariate Probit* model (Maddala [1983]).

Wilde (2000) shows that, provided that a varying exogenous regressor is present in each equation, exclusionary restrictions are not needed in order to ensure identification. In that case, however, one would heavily rely on distributional assumptions. In order to avoid identification by functional form, I follow the suggestion by Monfardini and Radice (2008) and I include, in the vector X_M, a set of variables that are exogenous to the individual labor supply decision.

3.3.1 Remittances and Labor Supply

The baseline equation to be estimated is:

$$L = \alpha'_X X_L + \alpha_R R^* + \epsilon_L \tag{4}$$

where L indicates the labor outcome (at either the extensive or intensive margin) and R^* represents the amount of remittances received by the respondent. We want to obtain a consistent estimate of the parameter α_R, that measures the impact of remittances on labor supply.

The task is quite challenging, since the estimation is complicated by two problems. The first problem is that not all the respondents receive remittances from abroad. A necessary condition for observing the variable R^* is the presence of an international migrant among the family members. This means that R^* is a (left) censored variable.

[73] With $E(\epsilon_L) = E(\epsilon_M) = 0$, $\text{Var}(\epsilon_L) = \text{Var}(\epsilon_M) = 1$ and $\text{Cov}(\epsilon_L, \epsilon_M) = \rho$.

Specifically:

$$R^* = \begin{cases} \theta'_X X_R + \epsilon_R, & \text{if } M^* > 0 \\ \text{unobserved}, & \text{otherwise} \end{cases} \tag{5}$$

One has therefore to take into account the presence of self-selection in the pool of *migrant* households (and, hence, in the pool of potential remittances' receivers). Moreover, as discussed previously, this selection process is likely to be non-random and affected by unobservables (u) correlated with the error term ϵ_L in Equation 4. Also in this case, it is not possible to define *a priori* the direction of the bias introduced by the endogeneity of the selection process, since that depends on how ϵ_M in Equation 3 and ϵ_L in Equation 4 are correlated with the common unobservables u.

The second problem concerns, instead, the possible simultaneity between remittances' receipts and labor outcome, so that $E(R^*\epsilon_L) \neq 0$. Shocks that affect the respondent's labor supply may cause consequential changes in the remittances' receipts, so that the direction of causality is reversed. As an example, suppose that an individual living in Albania has a son who migrated to Italy. Assume moreover that this individual is hit by a negative shock (unobserved by the econometrician) that reduces his labor supply (suppose, as an example, that he is fired so his labor supply is now zero). Because of this negative shock, it is likely that he will demand more remittances from his son in Italy. As a consequence we would still see a negative relationship between labor supply and remittances' receipts in Equation 4 but, in this case, the direction of causality would go from L to R^* and not vice versa.

In order to tackle these two problems, I follow a multiple-step estimation strategy proposed by Haveman et al. (1991)[74] and applied also in Heim (2007) for the estimation of married women's wage and income labor supply elasticities.

[74]They analyze a problem of similar nature. Their objective is to estimate the effect of disability transfers on the labor supply decisions of the elderly population. As in my case, they too have a problem of non-random endogenous self-selection in the sense that in order to receive a disability transfer, an individual has to apply to a disability program and it is possible that unobservables that drive the decision of whether to apply or not, affect also the individual propensity to work.

I begin with estimating a selection process *a la* Heckman (1979). Under the assumption of ϵ_M and ϵ_R being jointly normally distributed, it is possible to obtain a consistent estimate of the θ_X parameters in Equation 5 by using a two-step procedure.

The first step consists of estimating a Probit equation that describes the selection of the sample into the *migrant* (and therefore potential remittances' receivers) category. After obtaining consistent estimates of β_X, one can use the fitted values to compute the inverse Mills ratio

$$\hat{\lambda} = \frac{\phi(\hat{\beta}'_X X_M)}{\Phi(\hat{\beta}'_X X_M)}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the density and the cumulative distribution function.

In the second step, I estimate Equation 5 including, among the regressors a set of variables $\mathbf{Z}_M$ correlated with the remittances receipts but not with the individual labor supply. The inclusion of this vector will correct for the reverse causality of remittances. Moreover the inverse Mills ratio $\hat{\lambda}$ is added to control for the self-selection into the sample of migrant households. This procedure produces consistent estimates of θ_X and θ_λ (where the latter is the coefficient associated to the $\hat{\lambda}$ term in the remittances equation). With the estimated coefficients I can compute the residuals $\hat{\epsilon}_R$ and use a Control Function approach to estimate the following equation

$$L = \alpha'_X X_L + \alpha_R R + \gamma \hat{\epsilon}_R + \epsilon_L \tag{6}$$

to obtain an unbiased and consistent estimate of α_R, the effect of remittances on labor supply.

3.4 Data

Albania represents an interesting case study for the analysis of the impact of international migration.

During the 1990s Albania went through a dramatic migration process that started with the collapse of the communist regime in 1991 and the following economic crisis. According to the 2001 Albanian Census, between 1989 and 2001, about 800,000 people have migrated out of Albania, the largest emigration movement in Europe after World War II. Scenes, like the one presented in Figure 1, were not unusual in the ports of South Italy in the months between 1991 and 1993.

The rapid transition from central planning to a market economy was not accompanied by a reform of the financial system. The deficiency of the financial sector led to the development of informal deposit-taking companies that based their success on the creation of pyramid schemes. The system was obviously destined to collapse, as exactly happened in November 1996. When the schemes collapsed there was unconstrained rioting, the government fell and the country descended into anarchy and a near civil war.[75] This fueled an additional outflow of people from the country; according to estimates by the World Bank, in 2005, 860,000 Albanians still lived abroad (mainly in Greece, Italy, the U.K. and Germany) representing about 28% of total population.[76]

Figure 2 shows the absolute number of international migrants from each of the 36 Albanian administrative districts during the 1989-2001 period. Most of the emigration flows originated from the coastal area (mainly due to its vicinity to Italy) but about all the districts have been characterized, to some degree, by an outflow of the resident population.

[75] It is estimated that, at their peak, the nominal value of the pyramid schemes' liabilities amounted to almost half of the country's GDP and about two-third of Albanian population had invested in them.

[76] World Bank, Development Prospects Group (2005), *Migration and Remittances Factbook*.

The increasing number of people migrating abroad was followed by an increasing amount of remittances flowing to Albania. Figure 3 reports the absolute amount (in constant 2000 U.S.$) of remittances between 1995 and 2010, as well as their share with respect to GDP.[77] As we can see, the value of remittances receipts constantly increased during the early years of 2000s, experienced a drop in 2007 in correspondence with the beginning of the global economic crisis, and it amounted to about U.S.$600 million at the end of 2010. In terms of their share to GDP, on average, remittances account for about 10 percent of the Albanian domestic product, making them the second most important source of external financing after Foreign Direct Investments.[78]

The main data source for this paper is the 2005 Albania Living Standard Measurement Survey (LSMS05). The survey contains information at individual and household level for a sample of 3,840 households (17,302 individuals) from 455 census enumeration areas, drawn with the use of a multi-stage cluster design. The LSMS05 includes a typical questionnaire covering general household demographics, educational levels, asset ownership and labor market participation. Moreover the survey contains also community-level data, with information on the local access to services and infrastructure. The interesting feature of LSMS05 is that it provides an unusually rich section on migration, with detailed information on the household's network abroad.

In order to define labor participation, I focus only on the working age individuals (20-60 years of age), excluding those who are not in the labor force because of chronic illness or disability, those who are in military service, those who are classified as full time students and those individuals that have retired. The survey allows us to measure also the intensive margin of the individual labor supply since, for each worker (as defined above), it gives an indication of the number of hours per week that he/she

[77] Source: Word Economic Indicators, The World Bank.

[78] Albania ranks 18th among the top remittances recipient countries, in terms of remittances-to-GDP ratio (Source: *Migration and Remittances Factbook 2011*, The World Bank).

usually works.[79] This gives me a final sample of 2,193 male heads of households and 2,505 females that are either heads of household or partners of the head of household. These two groups will represent the samples at the core of my analysis.

Table 1 provides a concise description of the exogenous regressors used in the empirical analysis. As common in the empirical labor literature, I control for the usual individual and household characteristics such as: age (and its squared), marital status, human capital (through a series of dummies indicating the highest level of education achieved), family structure (household size and its square, the number of male and female adults, of children and of members still in school). The economic situation of the household can have an impact on the individual decision of whether or not participating into the labor market. However, the introduction of a measure of income among the regressors would likely introduce some endogeneity issues. Therefore I control for the household wealth level by relying on two variables indicating the subjective adequacy of general and food expenses, and on a set of dummies measuring the household ownership of some assets. Finally, in order to take into account the social and economic conditions at district level, I also include a measure of income inequality (Gini), the district level of unemployment rate and the share of population with secondary education.[80]

Table 2 presents some descriptive statistics for the males' and females' sample. In both samples I define two groups: individuals living in a household with (at least) an international migrant ($Migrant = 1$) and those living in a household without any member residing abroad ($Migrant = 0$). As we can see from the table, the two groups are quite different from each other with respect to several dimensions. *Non-migrant*

[79] If an individual works less than 40 hours/week, he/she is asked the reason. If that is due to illness or disability of the worker, or because the worker is still in school, I excluded him/her from the sample of workers.

[80] These last three variables are taken from the 2001 Albanian Population and Housing Census (PHC 2001).

households have a higher number of members.[81] Moreover, as expected, females are more likely to be the head of *migrant* households since, in some cases, the husband is the one who migrated abroad. The level of education appears to be higher in the *non-migrant* households which are characterized by a higher percentage of male heads or female heads/partners with secondary or college education.

These differences are confirmed also when one looks at variables that indicate the work status of the household members. The Table focuses on two indicators that will also be at the center of the econometric analysis: labor force participation (extensive margin) and number of hours usually worked during the week (intensive margin). Focusing first on the extensive margin, as expected, the labor force participation is on average higher for males compared to females. Within the two samples, though, the participation rate differs according to whether the individual has a family member living abroad or not. The labor participation rate is lower for subjects (either males or females) living in *migrant* households and the difference is highly statistically significant in both samples. The same pattern is found also looking at the intensive margin of the labor supply: the number of hours usually worked during the week is significantly higher for individuals living in a *non-migrant* household relative to those living in a *migrant* household. One possible explanation for these results is that members of *migrant* households can rely on a source of non-labor income (remittances) and therefore can rationally decide to decrease their labor supply. The data suggest that a significant share (60 percent for the males' sample and 66 percent for the females' sample) of households with members abroad have received remittances in the last twelve months.

The survey provides, among other things, a significant amount of information about the migrants. Table 3 summarizes some of the available information about them. There are 2067 migrants, belonging to 1184 households. Confirming the gen-

[81] Here I have considered only members that live in the household. Therefore migrants are not included in the count.

der bias in the Albanian migration process, the majority (65 percent) of the migrants are male. Moreover, while it is common for males to migrate individually, usually females migrate with their respective families: 89 percent of the female migrants live abroad with their partner as opposed to only 39 percent for male migrants. Two additional points are worth being stressed. First, female migrants tend to be slightly better educated than their male counterparts (14 percent of women have some degree of tertiary education, while only 6 percent of men do). Second, male migrants show, on average, a propensity to remit (67% versus 33%) higher than that of females.

3.5 The Instruments

The aim of this section is to provide a description of the variables used as instruments in Equation 3 (the migration equation) and in Equation 5 (the remittances' receipts equation).

3.5.1 Instruments for Migration

Valid instruments have to satisfy two conditions: they have to be (i) correlated with the decision of sending a family member abroad and (ii) uncorrelated with the labor supply decisions of those left behind.

I identify the following variables as potential instruments: the historical migration rate of the district where the household resides, and the gender composition of the household.

The *District Migration Rate* is an indicator of the rate of emigration (at district level) between 1990 and 2001. This variable wants to capture the network effect of international migration. Intuitively, if a household lives in a district with a high historical emigration rate, it will more likely send a family member abroad (relative to a household residing in a different district with a lower historical migration rate) because it can count on a network of people coming from the same geographical re-

gion that can ease the prospective migrant's adaptation to a foreign country. The first reason why I claim the exogeneity of this variable with respect to the individual labor supply decisions, is that the emigration rate is computed over a period of time that does not overlap with the survey's year (2005). However, it is possible that districts with a higher emigration rates are also characterized by higher unemployment rates or other unfavorable economic conditions somehow persistent over time. In that case, the presence of correlation between the instrument and individual labor outcome can't be excluded. In order to respond to this criticism I also include, as controls, variables that capture the district-level economic and social conditions (specifically the unemployment rate, an indicator of economic inequality and an indicator of the average level of education in each district). Conditioning on those variables, I argue that the only effect that the 1990-2001 district-level emigration rate can have on the individual labor outcome (in 2005) is via migration and not via other (potentially confounding) channels.

The Albanian society is still mainly patriarchal. Women are typically confined inside the house, where their main activities are related to taking care of the domestic chores and the household's members (especially children and the elderly). Due to this view of their role in the society, Albanian women are still characterized by a low labor force participation and face higher unemployment rates than their male counterparts.[82] Therefore, in an environment like the Albanian one, female and male labor are far from being perfect substitutes. Consider now a simple family composed by father, mother, a son and a daughter. In such familial context, the cost (in terms of labor loss) associated to the emigration of the only male son may be larger than its potential benefit.[83] If instead that family had an additional son, it would then

[82]According to estimates by the World Bank, the adult female labor force participation in 2004 was equal to 55 percent, compared to the adult male labor force participation of around 80 percent.

[83]It is also quite unlikely that the daughter would migrate. The Albanian migration process is primarily male-oriented. According to INSTAT (2001), about 75 percent of the migrant population is male. The majority of women migrate as part of a family, either together with the family or following their husband/partner after he had established himself in the host country. Single women

be easier to substitute the labor loss associated with the departure of one son. It is therefore likely that the opportunity costs associated with the migration choice decrease with the household's number of male children. To take into account this fact, I include among the instruments a set of dummy variables that indicate whether the extended household (that is those living in Albania plus those migrated abroad) has only one, two, three or more than three male members ($Male_1$, $Male_2$, $Male_3$ and $Male_{3plus}$). The intuition is that, conditioning on the household size, the larger the number of males in the extended household, the more likely that same household will send a member abroad since the cost associated to the loss of labor will be smaller.[84]

3.5.2 Instruments for Remittances

In order to correct for the reverse causality between remittances receipts and individual labor outcomes, I need variables that are correlated with the former but uncorrelated with the latter.

Exploiting the variation of variables at the destination country level seems to represent a convenient way to satisfy those two conditions. In particular I am using the following instruments: *Partner*, *Children*, *International GDP* and *Time Away*.

The variable *Partner* indicates the household share of migrants living abroad with their partner, while the variable *Children* measures the share of migrants that live abroad with their children. These variables are likely to be correlated with the frequency and/or the amount of remittances that the migrants send back to Albania (since, *ceteris paribus*, the amount of resources that can be sent back home will decrease) and are uncorrelated with the individual supply choices of the family members

migrating abroad still represent an exception.

[84] Alternatively I could have used a continuous variable instead of the set of dummies. That strategy, though, would not reveal the presence of potential non-linearities. When the number of male children is entered linearly variable is used, the results are qualitatively and quantitatively equivalent.

left behind.

The variable *Time Away* measures the average number of consecutive years that the migrants have spent abroad since they left the country. The reason for including this variable (and its squared term) in the remittances' equation is that remittances from Albanian migrants seem to be consistent with a sort of *decay* behavior. Given that this variable is positively related to the age of the family members, and since the latter influences the individual labor supply decisions, it is necessary to control for the age of the respondent to preserve its exogeneity. Finally, *International GDP* (GDP^W) is a variable based on the (weighted) real GDP per capita of the locations where the migrants live. The survey contains very detailed information about the international migrants. In particular, it is possible to know the country and the province where they live. Using this information, I matched each location with GDP per capita data taken from different sources: Eurostat (at territorial level 3) for European countries (which represent more than 90 percent of the Albanian migrants' destinations) and equivalent national statistical agencies for the remaining extra-european countries.[85] The variable is then built as:

$$GDP^W_{i,c} = \sum_j \omega_{i,c,j} GDP_j$$

where GDP_j is the 2004 real GDP per capita of the province j, weighted by $\omega_{i,c,j}$, which is equal to the share of migrants, originating from the municipality c that the household i belong to, living in the foreign province j.[86]

The reason for including this variable is that it is likely to positively affect the amount of remittances that migrants send back to Albania. At the same time, though, this variable does not affect the labor market participation of the individuals left in

[85]Ideally, in order to convert GDP data in real terms, I would need the price level in each province. I don't have such information. Therefore I have used the GDP deflator at country level. The assumption is that, within the same country, the relative prices do not differ across provinces.

[86]In the data I can identify 204 municipalities.

Albania, since it is linked with factors that are measured at the destination countries level.

3.6 Results

3.6.1 International Migration and Labor Supply

I begin by presenting the results concerning the analysis of the relationship between labor supply and international migration. Since having a family member living abroad is a prerequisite to the receipt of international remittances, these results can give us an idea of what to expect from the study of the labor supply-remittances nexus.

Table 4 presents the full information maximum likelihood estimates of the bivariate probit model represented by Equation 2 and Equation 3. Columns II and IV report the estimated coefficients for the migration equation. The instrumental variables have the expected signs. Specifically, living in a district with a high historical emigration rate increases the probability of having a migrant among the family members. Moreover, as the number of male members in the extended family increases, the higher is the likelihood of belonging to a *migrant* household. Columns III and VII present, instead, the recursive bivariate Probit's estimated effect of the variable M on the respondent's labor force participation. Based on these estimates, having a family member living abroad reduces the individual labor force participation by 13 percent (for a male head of household) and by 23 percent (for a female head of household or partner of the head of household).[87]

We can compare these results with what we would obtain if we estimated Equation 2 with a simple univariate Probit without taking into account the endogeneity of the self-selection into migration. Results are reported in columns I and V. Not

[87] For a discussion about the computation of the marginal effects after a recursive bivariate Probit, see Greene (1998).

correcting for the endogeneity of M tends to generate an upward biased estimated coefficient (i.e. a smaller negative effect of migration on the dependent variable).

The source of the bias is represented by the correlation, ρ between the two error terms ϵ_M and ϵ_L, for which the Table reports an estimate. The two error terms appear to be positively correlated (and this correlation is statistically different from zero), suggesting that unobserved factors that increase the probability of having a family member abroad have a positive effect also on the individual probability of participating in the labor market.[88]

Given the significant correlation between the error terms in the labor participation and migration equation, the estimation of a causal impact of M relies on the availability of appropriate exogenous regressors that are: *(i)* correlated with the endogenous regressor M; *(ii)* uncorrelated with the error term ϵ_L in Equation 2.

It is easy to show that our instruments meet the first test. Table 4 reports a χ^2-test for the null hypothesis of the effect of the instrumental variables on $\Pr(M = 1)$ being jointly equal to zero. As we can see we can always comfortably reject the null hypothesis. In order to prove the second point, I follow the suggestion of Evans and Schwab (1995) and I perform a single-probit estimation of Equation 2 where, among the regressors, I include also the variables used as instruments. This does not represent a formal test, since, given that the correct specification is a bivariate probit, then the single equation model is misspecified. However, the statistical non-significance of the coefficients attached to the instruments would suggest that they do not have any effect on the dependent variable (L) other than the one through the endogenous variable M. Results, reported in columns V and X, reveal that this is exactly the case.[89]

[88] In the last row of Table 4 a Wald test for $\rho = 0$ is reported. The null hypothesis of absence of correlation between ϵ_M and ϵ_L can be comfortably rejected.

[89] Formal tests of weak instruments and overidentifying restriction in case of a binary model with a dichotomous endogenous regressor are not currently available. Nichols (2011) suggests that the tests performed through a linear model (i.e. IV-2SLS) represent a useful starting point. If I follow that suggestion, I find that, for both the sample of males and females, the instruments always comfortably

In order to test the robustness of the previous results, I use, as an alternative strategy, the *two-stage residuals inclusion* (2SRI) estimation. This method produces consistent estimates in non-linear models with a dummy endogenous variables, as described in Terza et al. (2008). First, I estimate an auxiliary regression where the dependent variable is the endogenous regressor M. This consists, basically, of estimating again the *migration equation* of the recursive bivariate Probit.[90] Then, following Gourieroux et al. (1987), the generalized residuals are calculated as:

$$\tilde{\epsilon} = \frac{\phi(\widehat{\Gamma})}{\Phi(\widehat{\Gamma})[1 - \Phi(\widehat{\Gamma})]}[M_{i,h} - \Phi(\widehat{\Gamma})] \tag{7}$$

where $\widehat{\Gamma} = \hat{\beta}'_X X_M$.[91] The generalized residual vector $\tilde{\epsilon}$ is then included as an additional regressor in Equation 2 to capture the effect of unobservables that influence both the endogenous regressors and the labor force participation decision.

Columns IV and VIII in Table 4 show the 2SRI estimates. After correcting for endogeneity through the inclusion of the generalized residuals, the coefficient on the migration variable is negative, statistically significant and very close to the value obtained with the bivariate Probit model. Moreover the estimated coefficient on the generalized residuals is positive and significant, suggesting, once again, that selection into migration depends on unobservables that increase the probability of participating in the labor market.

Using the same generalized residuals previously obtained, it is possible to estimate the impact of international migration on the intensive margin of the labor supply (i.e. now the dependent variable in Equation 2 is the number of hours that the respondent usually works during the week). Table 5 presents the main results without (columns I and III) and with (columns II and IV) the inclusion of the generalized residuals. As

pass the weak identification test (Kleibergen-Paap F statistic larger than 10) and the Hansen test of overidentifying restrictions.

[90] The estimated coefficients are essentially the same shown in Table 4, columns II and VI

[91] $\Phi(\cdot)$ and $\phi(\cdot)$ represent, respectively, the cumulative density function and the probability density function of the normal distribution.

we can see, both for the sample of males and the sample of females, not correcting for the endogeneity of the variable M leads to an underestimation (in absolute terms) of the impact of international migration on the intensive margin of labor supply.

The correct estimated marginal effects (conditional on labor force participation) suggest that having a family member abroad reduces the weekly labor supply by about 7 hours for males and by about 3.5 hours for females. Notice moreover that the sign attached to the generalized residuals' term is positive, indicating, once again, that unobservables that affect positively the selection into migration, have also a positive impact on the amount of labor that the respondent supplies in the labor market.

The finding that the presence of a migrant in the family has a stronger effect on women at the extensive margin and on men at the intensive margin, may be indicating that females (either head of household or partner of the head of household), rather than reducing the number of hours supplied, tend to exit directly the labor market in response to the migration of a family member.

3.6.2 Remittances and Labor Supply

In order to obtain a correct estimate of the effect of the remittances' receipts on the labor supply of the members left behind (α_R in Equation 4), I use a three-step estimation strategy.

The first step consists of estimating Equation 3 which will account for the selection into migration (and therefore into the pool of remittances' receivers). The estimated coefficients, for both the males and females sample, are essentially the same as the *migration equation* in the recursive bivariate Probit, reported in Table 4, columns II and VI.

In the second step, on the relevant sample of individuals with non-missing remittances (i.e. individuals belonging to *migrant* households), I perform an OLS regression of Equation 5 where the dependent variable is represented by the logarithm of remit-

tances received. Among the regressors I included also an additional variable, $\hat{\lambda}$, which represents the inverse Mills ratio computed using the fitted values from the previous step and that corrects for the non-random self-selection into migration (i.e. into the pool of remittances' potential receivers).

Results are shown in Table 6.[92] Focusing just on the variables used as instruments in order to solve for the reverse causality between remittances and labor supply, their effect on the (logarithm of) remittances receipts is as expected. The higher the share of migrants living with their partner or with their children, the lower the amount of remittances received. The *decay* behavior of remittances is confirmed since the coefficient on the variable *Time Away* is positive while the coefficient on its squared term is negative. An increase in the weighted GDP variable has, as expected, a positive effect on the amount of remittances received, indicating that more favorable economic conditions in the countries/provinces where the migrants live generate an increase in the flow of money sent back to Albania. Finally the coefficient attached to the inverse Mills ratio λ is significant for both samples, indicating that sample selection correction is needed.

The third step consists of deriving the series of predicted remittances, $\hat{R}$, using the coefficients just estimated in the previous step, and use it in place of R^* in Equation 4. This series represents the amount of expected remittances receipts after correcting for both the reverse causality problem and the endogenous self-selection into migration. Table 7 reports the estimated coefficients on the extensive margin, while Table 8 presents the results concerning the intensive margin.[93]

Table 7 reports also the elasticities of labor force participation with respect to

[92]Note that the standard errors for this *outcome* equation have to be corrected for two reasons. The first is that we have heteroskedasticity if $\gamma_\lambda \neq 0$ (i.e. if we have selection bias). This is corrected by using robust standard errors. The second reason is that the inverse Mills ratio, derived from the *selection* equation, is estimated with uncertainty. In order to take this fact into account,the standard errors have been corrected via bootstrap (500 replications).

[93]The reported standard errors have been bootstrapped in order to take into account the inclusion of $\hat{R}$ among the regressors.

remittances receipts. Remittances have a particularly strong impact on the female labor force participation. According to the estimates, a 1 percent increase in the annual remittances receipts determines a 0.05 percent decrease in the labor force participation for females and a 0.015 percent decrease for males.[94] The estimated remittances-labor participation elasticities lie within the range of the estimates found in the labor literature (see, for example, Blomquist [1983], MaCurdy *et al.* [1990], Van Soest *et al.* [1990], Bourguignon and Magnac [1990], Flood and MaCurdy [1992], Heim [2007]).

Notice, moreover, that not accounting for the endogenous self-selection of the migration process and the reverse causality of remittances, would introduce an upward bias in the estimated coefficient (columns I and III), particularly for the female sample. As a specification check, Table 6 reports also the results from tests for weak instruments, including the partial R-squared of the instruments, as well as the F-statistic and the related p-value from the test that the coefficients on all instruments are zero. Staiger and Stock (1997) suggest that F-statistics above 10 are preferable in order to avoid problems related to weak instruments. As we can see from the reported results, the F-statistic is above that threshold.

Following Amuedo-Dorantes and Pozo (2006) I estimated Equation 4 using a Tobit model in order to account for the zero-inflated nature of the dependent variable.

Table 8 reports both the estimated coefficients and the marginal effects conditional on labor force participation. The results are as expected: an increase in the amount of remittances received during the year determines a decrease in the number of weekly hours worked. At the margin the effect appears to be stronger for males compared to females (and this is in line with the results found in Table 5): a one percent increase in remittances diminishes the weekly labor supply by 0.5 hours for males and 0.4 hours for females. Notice, however, that the effect, once correcting for

[94]Remittances participation elasticities are calculated as $\alpha_R \left[\frac{\phi(\hat{\beta}'_X X_M)}{\Phi(\hat{\beta}'_X X_M)}\right]$

the endogeneity of remittances, is never statistically significant, suggesting that the impact of remittances on labor supply occurs at the extensive margin rather than at the intensive margin.

3.6.3 Further Results

The results shown so far present an average estimated effect. It is now interesting to consider the same estimation model presented above on different subsamples of the population in order to explore if the impact of remittances on labor supply is differently shaped according to individual attributes such as the place where agents live (rural vs. urban), their age and their educational level. It is plausible that some degree of heterogeneity occurs given that labor market returns and conditions may differ across contexts and individual characteristics.

Table 9 (columns I and II) shows that, at the extensive margin, the negative effect of remittances is mainly concentrated in the 41-60 age group for the males' sample (the estimated marginal effect for the younger cohort is actually positive but not statistically significant), while it is almost evenly contributed by the two age groups (with a relatively higher effect in the 20-40 cohort) in the females' sample. A similar pattern (columns III and IV) is found as concerns the intensive margin of labor supply, but the only (mildly) significant impact of remittances is encountered in the older cohort of the male heads of households.

In Table 10 individuals are differentiated according to their level of education: primary (or no education), secondary (including vocational) education and college education. The results are quite different when we compare the males' and females' sample. For male heads of household, at the extensive margin, the negative impact of remittances is concentrated among the individuals with primary education and among those with college education. College educated men, moreover, appear to re-

spond more strongly to the receipt of international remittances. A more "canonical" pattern is instead experienced by the females sample. In this case, indeed, remittances display a monotonic impact on labor force participation, with less educated women responding more strongly to the receipt of money from family members living abroad. Remittances determine a decrease in the weekly number of hours worked for lower and college educated men (but the effect is statistically significant only for the first group) while they display the same pattern seen at the extensive margin for the sample of women (but the estimated effect is never significant).

Labor market conditions can be quite different between rural and urban areas. Consequently I have examine the presence of heterogeneity based on the location where the individuals live, differentiating between rural and urban environments. Results are presented in Table 11. In this case the effect of remittances, at both the extensive and intensive margin, show a common pattern: for both males and females, individual labor supply responds more strongly in urban areas. One explanation for this finding may be that under-employment (in the form, for example, of unpaid work in the family farm) is more common in a agricultural environment so that, even if an individual leaves the status of salaried worker, she still continues to supply labor in order to meet the needs of the family farm/rural enterprise.

One potential concern with the results presented so far is that they abstract from the possibility of labor redistribution within the household. It is possible, indeed, that remittances, as suggested by the empirical evidence contained in the previous sections, determine a decrease in the labor supply of male heads and female partners/heads of households, and, at the same time, an increase in the labor supply of the other members of the household so that we only experience a redistribution of labor within the family and not, as claimed so far, a reduction of the labor effort. In order to test this point, I have to change the observational unit from a single individ-

ual (either the head of household or the partner) to the household itself.[95]

In order to evaluate the labor supply at household level, I define two new variables: *share employed members* and *average hours worked*. The first variable is the ratio between the number of workers in the household (i.e. the number of members who usually work at least one hour during the week) and the number of *potential* workers in the household. In order to define a family member as a *potential* worker, I adopt the same criteria previously defined: between 20-60 year old, not full-time student, not affected by chronic illness/disability, not in military service, not retired. The variables *average hours worked* is simply the ratio between the number of hours usually worked during the week by the household members and the total number of *potential* workers in the household. I then re-estimate Equation 4 using these two indicators as dependent variables. The results seem to confirm that the previous findings at individual level are not the result of a mere redistribution of labor within the household. Remittances do decrease labor supply (both in terms of participation in the labor market and in terms of the number of hours supplied) also when we consider the household as observational unit.

3.7 Conclusions

This paper analyzes the relationship between international migration, remittances and the labor supply outcomes of the family members left behind in Albania, a country that have experienced a massive emigration process during the 1990s.

As concerns international migration, after controlling for the endogeneity of the selection into migration with respect to the individual labor supply choices, results suggest that having a family member living abroad reduces the probability of participating the labor market, particularly for women.

[95] Results available upon request.

Adopting a new approach, based on 3-step estimation strategy and the use of instrumental variables to correct for both the endogeneity of self-selection into migration (and therefore into the pool of remittances receivers) and the simultaneity between labor choices and remittances, the main finding is that an increase in the amount of resources received from abroad reduces the labor supply of the members left in Albania at both the extensive and the intensive margin. This effect appears to be stronger for women (relative to men), and for individuals living in urban areas (relative to residents in rural areas).

3.8 Tables

Table 1: **Variables' Description**

lnRem	ln(1 + remittances)
Age	age of the respondent, in years
Age^2	
Educ. Secondary	dummy=1 if maximum level of education is secondary education
Educ. Vocational	dummy=1 if maximum level of education is vocational education
Educ. College	dummy=1 if maximum level of education is college education
Head	dummy=1 if respondent is head of household. Notice that this is always the case for the sample of males.
Married	dummy=1 if respondent is married or lives with partner
No. Children	number of children between the age of 0 and 6 in the household (excluding migrants)
No. Members in School	number of full time students in the household (excluding migrants)
HH Size	adult equivalent household size (OECD scale)
$(HH\ Size)^2$	
Urban	dummy=1 if respondent lives in a urban area
Food Adequate	dummy=1 if respondent thinks that level of food expenditure in the household is adequate
Expenses Adequate	dummy=1 if respondent thinks that level of general expenditure in the household is adequate
Gini	district level Gini coefficient of income inequality (Source: PHC 2001)
Unempl. Rate	district level unemployment rate (Source: PHC 2001)
Distr.Education	district level share of population with secondary education (Source: PHC 2001)
Assets	dummy variables that indicate the ownership of: TV, refrigerator, washing machine, dishwasher, computer, car, satellite dish
Geog. Dummies	coast, north-east, central, south-east

Table 2: Descriptive Statistics

	Male Heads			Female Heads/Partners		
	Migrant = 0	Migrant = 1	Difference	Migrant = 0	Migrant = 1	Difference
Head of HH	1.00	1.00	-	0.04	0.11	***
Age	43.21	52.56	***	39.29	47.80	***
Muslim	0.81	0.76	**	0.81	0.76	***
Education						
Primary	0.41	0.51	***	0.52	0.65	***
Secondary	0.46	0.37	***	0.39	0.31	***
College	0.13	0.12		0.09	0.04	***
Size of HH	3.06	2.76	***	3.05	2.65	***
# Males >18	1.40	1.46	*	1.37	1.23	***
# Females >18	1.46	1.56	***	1.48	1.56	***
# Members 6-18	1.52	0.85	***	1.54	0.91	***
# Children 0-5	0.45	0.09	***	0.43	0.15	***
Urban	0.59	0.48	***	0.59	0.49	***
Employed	0.82	0.77	***	0.42	0.38	**
Hours Worked	39.9	36.5	***	23.2	23.9	
Receive Remitt.	-	0.60	-	-	0.66	-
Value Remitt.	-	U.S.$ 942.79	-	-	U.S.$ 1,081.11	-

Table 3: Migrants' Profiles

	Males	Females	Total
Age			
< 20	0.04	0.01	0.03
20-25	0.28	0.26	0.27
26-30	0.24	0.25	0.24
31-35	0.15	0.20	0.17
35-40	0.13	0.14	0.13
> 40	0.16	0.14	0.15
Education			
Primary	0.53	0.35	0.47
Secondary	0.27	0.38	0.31
College	0.06	0.14	0.09
Living with Partner	0.39	0.89	0.56
Living with Children			
Legal Residence	0.83	0.92	0.86
Currently Working	0.93	0.55	0.80
Destination Country			
Greece	0.39	0.43	0.40
Italy	0.41	0.35	0.39
Other Europe	0.14	0.12	0.13
North America	0.05	0.09	0.06
Other World	0.01	0.01	0.01
Sent Remittances	0.67	0.33	0.55
N	1357	710	2067

Table 4: Migration and Labor Supply Extensive Margin

	Male Heads					Female Heads/Partners				
	Probit	Rec. Biv. Probit		2SRI		Probit	Rec. Biv. Probit		2SRI	
		Migration Eq.	Labor Supply Eq.				Migration Eq.	Labor Supply Eq.		
	I	II	III	IV	V	VI	VII	VIII	XIX	X
Migrant	-0.21**		-0.49***	-0.55***	-0.20**	-0.25***		-0.83***	-0.86***	-0.19**
	(0.09)		(0.17)	(0.21)	(0.09)	(0.09)		(0.22)	(0.26)	(0.09)
	[-0.05]		**[-0.11]**	**[-0.12]**		**[-0.08]**		**[-0.26]**	**[-0.27]**	
District Migration Rate		0.02***			-0.000		0.03***			-0.008
		(0.006)			(0.005)		(0.006)			(0.006)
$Male_2$		0.49***			-0.14		0.47***			-0.07
		(0.13)			(0.17)		(0.10)			(0.08)
$Male_3$		0.78***			-0.13		0.76***			-0.13
		(0.14)			(0.14)		(0.12)			(0.09)
$Male_{3plus}$		1.57***			0.12		1.39***			-0.23
		(0.17)			(0.18)		(0.13)			(0.18)
$\tilde{\epsilon}$				0.22**					0.39***	
				(0.10)					(0.13)	
N	2,193	2,193	2,193	2,193	2,193	2,505	2,505	2,505	2,505	2,505
Geographical Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ρ			0.18**					0.39***		
			(0.09)					(0.14)		
Wald test (H_0: $\rho = 0$)			-3.80**					-7.26***		
χ^2 test (H_0: inst.=0) (p-value)			0.000					0.000		

Note: other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, Male members > 18, Female members > 18, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies and a constant term. For a detailed description of these variables, please see Table 1.
Robust, district-level clustered standard errors are shown in (parentheses). ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.
Marginal Effects, evaluated at the mean of the covariates and conditional on the number of hours being positive, are indicated in [parentheses].

Table 5: **Migration and Labor Supply Intensive Margin**

	Male Heads		Female Heads/Partners	
	I	II	III	IV
Migrant	-2.84*	-10.59***	-4.85**	-12.54**
	(1.68)	(3.40)	(2.34)	(4.71)
$\tilde{\epsilon}$		4.99***		4.91
		(1.88)		(3.12)
N	2,193	2,193	2,505	2,505
Geographical Dummies	Yes	Yes	Yes	Yes
$\frac{\partial E(hours\|hours>0)}{\partial Migrant}$	-2.23*	-8.32***	-2.25**	-5.82**
	(1.31)	(2.76)	(1.07)	(2.67)

Note: the dependent variable is the number of hours that the respondent usually works during the week.

Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, Male members > 18, Female members > 18, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies and a constant term. For a detailed description of these variables, please see Table 1.

Robust, district-level clustered standard errors are shown in (parentheses). ***, **, * mean statistical significance at, respectively, 1, 5 and 10 percent.

Table 6: **Remittances Receipts**

Dep. Var.: Logarithm of Remittances Received	I	II
Share - Partner	-0.011***	-0.010***
	(0.004)	(0.003)
Share - Children	-0.007*	-0.007**
	(0.004)	(0.003)
Time Away	0.20***	0.21***
	(0.07)	(0.05)
(Time Away)^2	-0.006**	-0.007***
	(0.003)	(0.002)
$ln(GDP^W)$	0.99**	0.93**
	(0.50)	(0.42)
λ	-2.03***	-1.53***
	(0.42)	(0.33)
Geographical Dummies	Yes	Yes
R^2	0.18	0.18
F-stat.	10.61	12.57

.

Note: The dependent variable is the logarithm of annual remittances receipts (in U.S.$). Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, Male members > 18, Female members > 18, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies. For a detailed description of these variables, please see Table 1.
The standard errors (shown in parentheses) have been bootstrapped (500 replications) to take into account the inclusion of λ among the regressors. ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.
F-stat refers to the F-statistics for the hypothesis of joint insignificance of Share - Partner, Share - Children, Time Away, (Time Away)^2, $ln(GDP^W)$

Table 7: **Remittances Receipt and Labor Supply Extensive Margin**

	Male Heads		Female Heads/Partners	
	I	II	III	IV
lnREM	-0.024	-0.037**	-0.034***	-0.046**
	(0.014)	(0.017)	(0.012)	(0.021)
	[-0.008]	**[-0.036]**	**[-0.038]**	**[-0.048]**
$\widehat{\epsilon_R}$		0.034		0.036
		(0.021)		(0.041)
N	2,193	2,193	2,458	2,458
Geographical Dummies	Yes	Yes	Yes	Yes
R^2	0.10	0.11	0.17	0.16

Note: estimated coefficients from a Probit regression. The dependent variables is a dummy equal to one if respondent participates the labor market, zero otherwise. Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, Male members > 18, Female members > 18, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies, a constant term. For a detailed description of these variables, please see Table 1.
Robust, district-level clustered standard errors are shown in (parentheses); moreover the standard errors in column II and IV have been bootstrapped to correct for the inclusion of $\widehat{\epsilon_R}$. ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.
Participation remittances elasticities, evaluated at the mean of the covariates, are shown in [**parentheses**].

Table 8: **Remittances Receipt and Labor Supply Intensive Margin**

	Male Heads		Female Heads/Partners	
	I	II	III	IV
lnREM	-0.49*	-0.52	-0.59*	-0.77
	(0.28)	(0.38)	(0.32)	(0.59)
	[-0.39]	**[-0.40]**	**[-0.27]**	**[-0.35]**
$\widehat{\epsilon_R}$		0.05		0.41
		(0.38)		(0.94)
N	2193	2193	2505	2505
Geographical Dummies	Yes	Yes	Yes	Yes

Note: estimated coefficients from a Tobit regression. The dependent variables is the usual number of hours worked by respondent during the week. Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, Male members > 18, Female members > 18, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies, a constant term. For a detailed description of these variables, please see Table 1.

Robust, district-level clustered standard errors are shown in (parentheses); moreover the standard errors in columns II and IV have been bootstrapped (500 replications) to correct for the inclusion of $\widehat{\epsilon_R}$. ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.

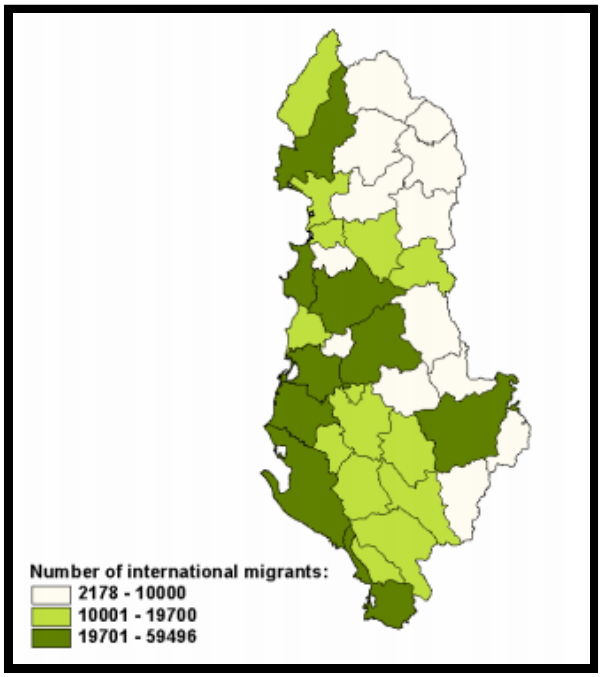
Hours-remittances marginal effects, evaluated at the mean of the covariates and conditional on the number of hours being positive, are shown in [**parentheses**].

3.9 Figures

Figure 1: March 7, 1991: 27,000 Albanians arrive in Brindisi (Italy).

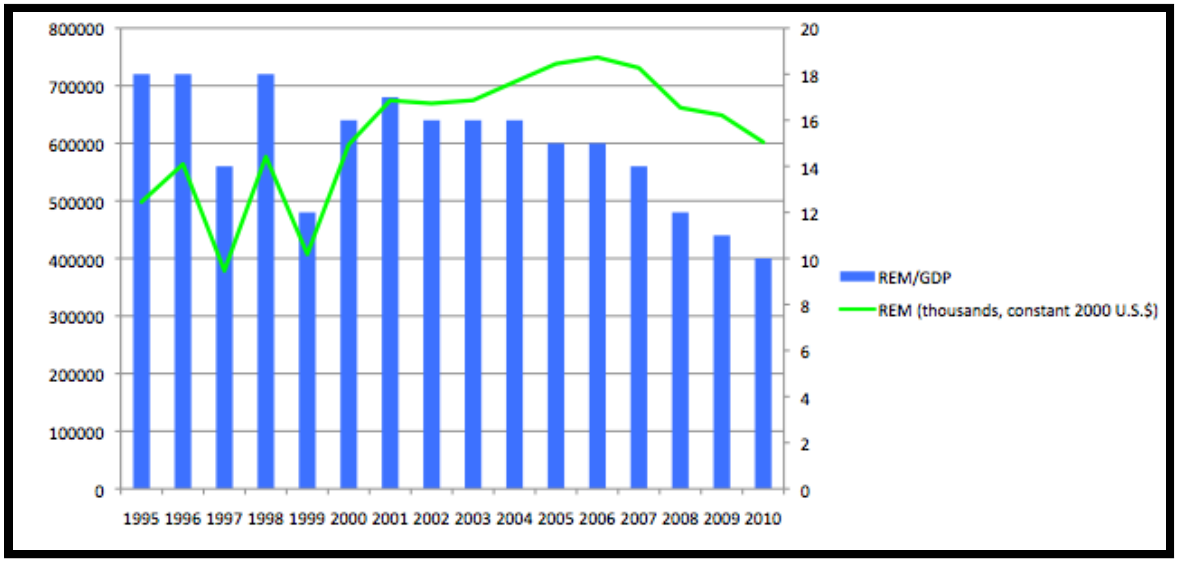
Source: *La Repubblica*

Figure 2: International Migrants by Districts, 1989-2001



Source: 1989 and 2001 Population and Housing Census

Figure 3: **Migrants' Remittances to Albania, 1995-2010**



Note: the figure displays the amount of real remittances ('000, in 2000 U.S.$) on the left axis and the ratio between remittances and GDP on the right axis. Author's calculations based on data taken from the World Development Indicators (World Bank).

3.10 Appendix A: Additional Tables

Table 9: **Heterogeneity: Age**

	Male Heads			
	Extensive Margin		Intensive Margin	
	20-40	41-60	20-40	41-60
$lnREM$	-0.071	-0.057***	1.27	-0.75**
	(0.064)	(0.021)	(1.88)	(0.31)
$\widehat{\epsilon_R}$	0.038	0.049**	0.76	0.29
	(0.068)	(0.025)	(2.83)	(0.46)
	Female Heads/Partners			
	Extensive Margin		Intensive Margin	
	20-40	41-60	20-40	41-60
$lnREM$	-0.042	-0.049**	-0.79	-0.48
	(0.031)	(0.020)	(0.74)	(0.56)
$\widehat{\epsilon_R}$	0.077	-0.018	1.02	-0.20
	(0.050)	(0.042)	(1.24)	(0.93)

Note: for the Extensive Margin Analysis, the dependent variables is a dummy equal to one if respondent participates the labor market, zero otherwise; for the Intensive Margin analysis the dependent variable is the respondent's usual number of hours worked during a week. Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies, a constant term. For a detailed description of these variables, please see Table 1.
District-level clustered, bootstrapped standard errors are shown in (parentheses). ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.

Table 10: **Heterogeneity: Education Level**

Male Heads

	Extensive Margin			Intensive Margin		
	Primary Ed.	Secondary Ed.	College Ed.	Primary Ed.	Secondary Ed.	College Ed.
$lnREM$	-0.069***	0.015	-0.29	-1.11**	0.33	-2.57
	(0.023)	(0.028)	(0.24)	(0.56)	(0.46)	(1.73)
$\widehat{\epsilon_R}$	0.058	0.002	0.062	0.93	-0.70	2.17
	(0.047)	(0.036)	(0.44)	(0.83)	(0.90)	(1.77)

Female Heads/Partners

	Extensive Margin			Intensive Margin		
	Primary Ed.	Secondary Ed.	College Ed.	Primary Ed.	Secondary Ed.	College Ed.
$lnREM$	-0.065**	-0.011	0.07	-1.17	-0.15	1.46
	(0.025)	(0.021)	(0.24)	(0.90)	(0.71)	(0.89)
$\widehat{\epsilon_R}$	0.056	-0.004	0.061	0.88	-0.37	-1.20
	(0.043)	(0.045)	(0.040)	(1.90)	(1.10)	(0.92)

Note: for the Extensive Margin Analysis, the dependent variables is a dummy equal to one if respondent participates the labor market, zero otherwise; for the Intensive Margin analysis the dependent variable is the respondent's usual number of hours worked during a week. Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Married, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies, a constant term. For a detailed description of these variables, please see Table 1. District-level clustered, bootstrapped standard errors are shown in (parentheses). ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.

Table 11: **Heterogeneity: Rural vs. Urban**

Male Heads

	Extensive Margin		Intensive Margin	
	Rural	Urban	Rural	Urban
$lnREM$	-0.05	-0.086**	-0.32	-1.31**
	(0.023)	(0.04)	(0.55)	(0.58)
$\widehat{\epsilon_R}$	0.048	0.051	0.59	0.41
	(0.043)	(0.04)	(0.96)	(0.67)

Female Heads/Partners

	Extensive Margin		Intensive Margin	
	Rural	Urban	Rural	Urban
$lnREM$	-0.024	-0.076***	0.15	-2.28***
	(0.028)	(0.026)	(0.62)	(0.76)
$\widehat{\epsilon_R}$	0.059	0.024	0.74	0.57
	(0.057)	(0.039)	(1.12)	(1.37)

Note: for the Extensive Margin Analysis, the dependent variables is a dummy equal to one if respondent participates the labor market, zero otherwise; for the Intensive Margin analysis the dependent variable is the respondent's usual number of hours worked during a week. Other variables included in the regressions and not shown (results can be provided upon request): Head of HH, Age, Age^2, Muslim, Secondary Ed., Vocational Ed. College Ed., Married, No. of Children, No. of Members still in school, HH Size, HH Size^2, Urban, Food Adequate, Expenses Adequate, District Gini, District Unemployment Rate, District Education, Assets Dummies, a constant term. For a detailed description of these variables, please see Table 1.
District-level clustered, bootstrapped standard errors are shown in (parentheses). ***, **, * mean, respectively, statistical significance at 1, 5 and 10%.

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.

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.

Acemoglu, D., S. Johnson, and J. A. Robinson (2005). The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth. *The American Economc Review 95*, 546–579.

Acosta, P. (2006). Labor Supply, School Attendance and Remittances from International Migration: the Case of El Salvador. World Bank Policy Research Working Paper Series.

Adams, R. H. (1992). The Impact of Migration and Remittances on Inequality in Rural Pakistan. *Pakistan Development Review 31*(4), 1189–1203.

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.

Amuedo-Dorantes, C. and S. Pozo (2006). Migration, Remittances and Male and Female Employment Patterns. *The American Economic Review: Papers and Proceedings 96*(2), 222–226.

ARHAP (2006). Appreciating Assets: The Contribution of Religion to Universal Access in Africa. Report for the world helath organization, ARHAP.

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

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.

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

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

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

Baur, J. (1994). *2000 Years of Christianity in Africa*. Nairobi, Kenya: Paulines Publications Africa.

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

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.

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

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

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

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.

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

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.

Blomquist, S. N. (1983). The Effect of Income Taxation on the Labour Supply of Married Men in Sweden. *Journal of Public Economics 22*(2), 169–197.

Bongaarts, J. (2007). Late Marriage and the HIV Epidemic in sub-Saharan Africa. *Population Studies 61*, 73–83.

Bourguignon, F. and T. Magnac (1990). Labor Supply and Taxation in France. *Journal of Human Resources 25*, 358–389.

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

Brockerhoff, M. and A. E. Biddlecom (1999). Migration, Sexual Behavior and the Risk of HIV in Kenya. *International Migration Review 33*, 833–856.

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.

Caldwell, J. (2000). Rethinking the Africa AIDS Epidemic. *Population and Development Review 26*, 117–135.

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

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.

Carletto, G., B. Davis, T. Kilic, and A. Zezza (2009). Investing Back Home: Return Migration and Business Ownership in Albania. World Bank Policy Research Working Paper No. 4366.

Carletto, G., B. Davis, J. Miluka, and A. Zezza (2010). The Vanishing Farms? The Impact of International Migration on Albanian Family Farming. *The Journal of Development Studies 46*(1), 140–161.

Catrinescu, N. (2009). Remittances, Institutions, and Economic Growth. *World Development 37*(1), 81–92.

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

Chami, R., S. Jahjah, and C. Fullenkamp (2003). Are Immigrant Remittance Flows a Source of Capital for Development? IMF Working Paper 03/189, International Monetary Fund.

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

Chiquiar, D. and G. Hanson (2005). International Migration, Self-selection and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy 113*(2), 239–281.

Comin, D., W. Easterly, and E. Gong (2010). Was the Wealth of Nations Determined in 1000 B.C.? *American Economic Journal: Macroeconomics 2*, 65–97.

Cox Edward, D. and M. Ureta (2003). International Migration, Remittances and Schooling: Evidence from El-Salvador. *Journal of Development Economics 72*(2), 429–461.

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.

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

Dabalen, A. and J. Miluka (2008). Exploring the Role of Albanian International Migration on Education. World Bank Working Paper.

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

Djankov, S., R. La Porta, and F. Lopez de Silanes (2003). Courts. *The Quarterly Journal of Economics 118*, 453–517.

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.

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

Dustmann, C. and O. Kirchkamp (2002). The Optimal Migration Duration and Activity Choice After Remigration. *Journal of Development Economics 67*(2), 351–372.

Engerman, S. L. and K. L. Sokoloff (1997). Factor Endowments, Institutions, and Differ- ential Paths of Growth Among New World Economies: a View from Economic Historians to the United States. In S. Harber (Ed.), *How Latin America Fell Behind*. Stanford, CA: Stanford University Press.

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.

Evans, N. E. and R. M. Schwab (1995). Finishing High Schools and Starting College: Do Catholic Schools Make a Difference? *The Quarterly Journal of Economics 110*(4), 941–974.

Flood, L. R. and T. MaCurdy (1992). Work Disincentive Effects of Taxes: an Empirical Analysis of Swedish Men. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 37, pp. 239–278.

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

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.

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

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

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.

Giuliano, P. and M. Ruiz-Arranz (2009). Remittances, Financial Development and Growth. *Journal of Development Economics 90*(1), 144–152.

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

Glaeser, E. L., R. La Porta, and F. Lopez de Silanes (2004). Do Institutions Cause Growth? *Journal of Economc Growth 9*, 271–303.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalized Residuals. *Journal of Econometrics 34*, 5–32.

Greene, W. H. (1998). Gender Economic Courses in Liberal Arts Colleges: Further Results. *The Journal of Economic Education 29*(4), 291–300.

Hallet, T. B., J. J. Lewis, C. A. Lopman, P. Mushati, M. Wombe, G. P. Garnett, and S. Gregson (2007). Age at First Sex and HIV Infection in Rural Zimbabwe. *Studies in Family Planning 38*, 1–20.

Hanson, G. and C. Woodruff (2002). Emigration and Educational Attainment in Mexico. mimeo, University of California at San Diego.

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.

Hardee, K., J. Gribble, S. Weber, T. Manchester, and M. Wood (2008). Reclaiming the ABCs: The Creation and Evolution of the ABC Approach. Population Action International Report.

Haveman, R., P. De Jong, and B. Wolfe (1991). Disability Transfers and the Work Decisions of Older Men. *The Quarterly Journal of Economics 106*(3), 931–959.

Hearst, N. and S. Chen (2004). Condom Promotion for AIDS Prevention in the Developing World: Is it Working? *Studies in Family Planning 35*, 39–47.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–162.

Heim, B. T. (2007). The Incredible Shrinking Elasticities: Married Female Labor Supply. *Journal of Human Resources 42*(4), 881–918.

Hoddinot, J. (1994). A Model of Migration and Remittances Applied to Western Kenya. *Oxford Economic Papers 46*(3), 459–476.

Iannacone, L. R. (1998). Introduction to the Economics of Religion. *Journal of Economic Literature 36*, 1465–1495.

INSTAT (2001). *Population and Housing Census of Albania*. Tirana, Albania.

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

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.

Johnson, H. B. (1967). The Location of Christian Missions in Africa. *Geographical Review 57*, 168–202.

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

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

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

Kohler, H.-P., J. R. Behrman, and S. C. Watkins (2007). Social Networks and HIV/AIDS Risk Perceptions. *Demography 44*, 1–33.

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

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. W. Vishny (1997). Legal Determinants of External Finance. *Journal of Finance 52*, 1131–1150.

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

Lopez-Cordoba, E. (2005). Globalization, Migration and Development: the Role of Mexican Migrant Remittances. *Economia 6*(1), 217–256.

Lucas, R. E. (1987). Emigration to South Africa's Mines. *The American Economic Review 77*(3), 313–330.

Lucas, R. E. and O. Stark (1985). Motivation to Remit: Evidence from Botswana. *Journal of Political Economy 93*(5), 901–918.

M., G., U. Larsen, and E. Paintsil (2009). Association of Age of First Sex with HIV-1, HSV-2 and Other Sexually Transmitted Infections Among Women in Northern Tanzania. *Transmitted Diseases 36*, 570–576.

MaCurdy, T., D. Green, and H. Paarsch (1990). Assessing Empirical Approaches for Analyzing Taxes and Labour Supply. *Journal of Human Resources 25*, 415–490.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, MA: Cambridge University Press.

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

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

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

McCarthy, N., G. Carletto, B. Davis, and I. Maltsoglou (2007). Assessing the Impact of Massive OutMigration on Agriculture. ESA Working Paper No. 06-14.

McCormick, B. and J. Wahba (2001). Overseas Work Experience, Savings and Entrepreneurship Amongst Return Migrants to LDCs. *Scottish Journal of Political Economy 48*(2), 164–178.

McCormick, B. and J. Wahba (2003). Return International Migration and Geographical Inequality: the Case of Egypt. *Journal of African Economies 12*(4), 500–532.

McKenzie, D. and H. Rapoport (2007). Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico. *Journal of Development Economics 84*(1), 1–24.

McKenzie, D. and H. Rapoport (2011). Can Migration Reduce Educational Attainment? Evidence from Mexico. *Journal of Population Economics 24*(4), 1331–1358.

Mesnard, A. (2004). Temporary Migration and Capital Market Imperfections. *Oxford Economic Papers 56*, 242–262.

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

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

Mohapatra, S., D. Ratha, and A. Silwal (2010). Outlook For Remittances Flows 2011-2012. Technical report, The World Bank, Washington D.C., U.S.A.

Monfardini, C. and R. Radice (2008). Testing Exogeneity in the Bivariate Probit model: Monte Carlo Evidence and an Application to Health Economics'. *Oxford Bulletin of Economics and Statistics 70*(2), 271–282.

Nichols, D. (2011). Causal Inference for Binary Regression.

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

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

Nunn, N. (2010). Religious Conversion in Colonial Africa. *The American Economc Review: Papers & Proceedings 100*, 147–152.

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

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

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

Oppong, C. (1995). A High Price to Pay: for Education, Subsistence, or a Place in the Job Market. *Health Transition Review Supplement to vol. 5*, 35–56.

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.

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.

Perez-Armendariz, C. and D. Crow (2010). Do Migrants Remit Democracy? International Migration, Political Beliefs and Behavior in Mexico. *Comparative Political Studies 43*(1), 119–148.

Pettifor, A. E., A. Van der Straten, M. S. Dunbar, S. Shiboski, and P. N. (18). Early Age of First Sex: a Risk Factor of HIV Infection Among Women in Zimbabwe. *AIDS 2004*, 1435–1442.

Pfutze, T. (2012). Does Migration Promote Democratization? Evidence from the Mexican Transition. *Journal of Comparative Economics 40*(2), 159–175.

Piracha, M. and F. Vadean (2009). Return Migration and Occupational Choice. IZA Discussion Papers 3922.

Potts, M., D. T. Halperin, D. Kirby, A. Swidler, E. Marseille, J. D. Klausner, N. Hearst, R. G. Wamai, J. G. Kahn, and J. Walsh (2008). Reassessing HIV Prevention. *Science 320*, 749–750.

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

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

Reniers, G. (2008). Marital Strategies for Regulating Exposure to HIV. *Demography 45*, 417–438.

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

Robinson, J. A. (1915). *History of Christian Missions*. New York, U.S.A.: Charles Scribner and Sons.

Rozelle, S., J. E. Taylor, and A. DeBrauw (1999). Migration, Remittances and Agricultural Productivity in China'. *The American Economic Review: Papers and Proceedings 89*(2), 287–291.

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

Schensul, S. L., A. Mekki-Berrada, B. K. Nastasi, R. Singh, J. A. Burleson, and M. Bojko (2006). Men's Extramarital Sex, Marital Relationships and Sexual Risk in Urban Poor Communities in India. *Journal of Urban Health*, 614–624.

Shelton, J. (2007). Ten Myths and One Truth About Generalised HIV Epidemics. *The Lancet 370*, 1809–1811.

Singh, R., M. Haacker, K. Lee, and M. Le Goff (2010). Determinants and Macroeconomic Impact of Remittances in Sub-Saharan Africa. *Journal of African Economics 20*(2), 312–340.

Smith, D. J. (2007). Modern Marriage, Men's Extramarital Sex and HIV Risk in Southeastern Nigeria. *American Journal of Public Health 97*, 997–1005.

Staiger, D. and J. H. Stock (1997). Instrumental Variables Regressions with Weak Instruments. *Econometrica 65*(3), 557–586.

Tabellini, G. (2007). Culture and Institutions: Economic Development in the Regions of Europe. mimeo.

Taylor, J. E. (1992). Remittances and Inequality Reconsidered: Direct, Indirect and Intertemporal Effects. *Journal of Policy Modeling 14*(2), 187–208.

Taylor, J. E. and T. J. Wyatt (1996). The Shadow Value of Migrant Remittances, Income and Inequality in a Household-farm Economy. *Journal of Development Studies 32*(6), 899–912.

Terza, J., A. Bazu, and P. Rathouz (2008). A Two Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of Health Economics 27*, 531–543.

Thompson, A. P. (1983). Extramarital Sex: A Review of the Research Literature. *The Journal of Sex Research 19*, 1–22.

UNAIDS (2007). AIDS Epidemic Update. Technical report, Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO).

UNAIDS (2010). Global report: Unaids report on the global aids epidemic 2010. Technical report, UNAIDS.

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.

Van Soest, A., I. Woettiez, and A. Kapteyn (1990). Labour Supply, Income Taxes and Hours Restrictions in the Netherlands. *Journal of Human Resources 25*, 517–558.

Wand, H. and G. Ramjee (2012). The Relationship Between Age of Coital Debut and HIV Seroprevalence Among Women in Durban, South Africa: a Cohort Study. *British Medical Journal 2*(000285).

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

WHO (2006). The World Health Report 2006: Working Together for Health. Technical report, WHO.

Wilde, J. (2000). Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors. *Economic Letters 69*, 309–312.

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

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

Woodruff, C. and R. Zenteno (2007). Remittances and Microenterprises in Mexico. *Journal of Development Economics 82*(2), 509–528.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Yang, D. (2008). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *Economic Journal 118*(528), 591–630.

Yates, D. (2003). Situational Analysis of the Church Response to HIV/AIDS in Namibia: Final Report. Technical report, Pan African Christian AIDS Network.