



**HAL**  
open science

**When the times they're not a changin': essays on the persistent effects of religion, investments, and ancestry on economic, social, and political behaviors at the subnational level**

Valeria Rueda

► **To cite this version:**

Valeria Rueda. When the times they're not a changin': essays on the persistent effects of religion, investments, and ancestry on economic, social, and political behaviors at the subnational level. Economics and Finance. Institut d'études politiques de paris - Sciences Po, 2016. English. NNT : 2016IEPP0027 . tel-03464543

**HAL Id: tel-03464543**

**<https://theses.hal.science/tel-03464543>**

Submitted on 14 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Institut d'études politiques de Paris  
ÉCOLE DOCTORALE DE SCIENCES PO  
Programme doctoral en économie

Département d'économie

Doctorat en Sciences économiques

When the times they're not a changin'

Essays on the persistent effects of religion, investments, and ancestry  
on economic, social, and political behaviors at the subnational level

Valeria RUEDA

*Supervised by:*

M. Yann Algan, Professeur des Universités, IEP de Paris

Defended in Paris, September 26<sup>th</sup>, 2016

*Jury:*

M. Yann ALGAN, Professeur des Universités, IEP de Paris

M. Sascha O. BECKER, Professor of Economics, University of Warwick (Rapporteur)

M. Sergei GURIEV, Professeur des Universités IEP de Paris

M. Nathan NUNN, Professor of Economics, Harvard University (Rapporteur)

M. Paul SEABRIGHT Professor of Economics, Toulouse School of Economics

Director of the IAST (Institute for Advanced Study in Toulouse)



Institut d'Études Politiques de Paris  
ÉCOLE DOCTORALE DE SCIENCES PO  
Programme doctoral en économie

Département d'économie

Doctorat en Sciences économiques

Les temps changent-ils?

Essais sur la persistance des effets de la religion, les investissements,  
et les origines culturelles, sur les comportements économiques et  
sociaux au niveau local

Valeria RUEDA

*Thèse dirigée par:*

M. Yann Algan, Professeur des Universités, IEP de Paris

Soutenue à Paris, le 26 Septembre 2016

*Jury:*

M. Yann ALGAN, Professeur des Universités, IEP de Paris

M. Sascha O. BECKER, Professor of Economics, University of Warwick (Rapporteur)

M. Sergei GURIEV, Professeur des Universités IEP de Paris

M. Nathan NUNN, Professor of Economics, Harvard University (Rapporteur)

M. Paul SEABRIGHT Professor of Economics, Toulouse School of Economics  
Director of the IAST (Institute for Advanced Study in Toulouse)



Pour Lucia, qui me montre la persévérance et qui m'a appris à nager



## Remerciements

Je souhaite remercier en premier lieu mon directeur, Yann Algan, de m'avoir ouvert les portes de Sciences Po, bien sûr, mais aussi de la LSE et de Cambridge. Je sais aussi que j'ai tendance à vouloir explorer toutes sortes de questions et le résultat n'a pas toujours été très convainquant du premier coup. Donc je remercie aussi sa patience. Yann m'a aussi introduite à l'équipe de CORE Economics, et m'a ainsi rassurée sur le fait qu'aimer enseigner avec des caricatures et des vidéos, certes parfois ridicules, ne fera pas de moi une moins bonne universitaire. Pour cela, je lui serai toujours reconnaissante.

Je tiens aussi à remercier Julia, ma co-autrice et amie, avec qui j'espère travailler pendant encore de nombreuses années. On ne pourrait pas être plus différentes, je pense que notre collaboration pourrait être un cas d'école de compétences complémentaires ! Personne ne m'a autant poussée et encouragée dans des moments de doute, ni remis sur le chemin de la recherche à tellement de reprises.

De nombreux professeurs m'ont inspirée et guidée dans mes recherches et dans ma vie de prof. Je veux prendre un moment, pour les saluer et les remercier. Leur enseignement reste avec moi à chaque fois que je prends à mon tour la parole devant des étudiants. De Brigitte Agard, Ghislaine Giol, et Philippe Pichon, j'ai retenu que l'enseignement avec rigueur de la rigueur est la meilleure manière de faire comprendre des concepts aussi abstraits soient-ils. Lydie Calame m'a inculqué dès un jeune âge un grand intérêt pour les questions de justice sociale. Beaucoup plus tard, Élise Huillery a répondu patiemment à mes questions sur le développement économique et m'a rassurée quant à la compatibilité totale de l'excellence avec l'humilité et l'honnête intellectuelle. De la même manière, je souhaite exprimer ma reconnaissance à Guy Laroque qu'une carrière prolifique et donc un emploi du temps sans doute très chargé, n'empêche pas de prendre le temps toutes les semaines, pour former les jeunes à la recherche et à sa critique, pendant les séances de groupe de lecture. Cet exercice m'a certainement permis d'aiguiser ma lecture des articles, et bien sûr, de critiquer ma propre recherche plus efficacement. Daniel Cohen sera, toujours, un grand exemple quant à sa capacité à ramener les questions les plus complexes aux mots les plus simples ; c'est un mentor extraordinaire pour toute personne qui souhaite enseigner l'économie ou transmettre ses messages dans l'espace public.



Nathan Nunn has been my role model since the moment I decided to change paths in my research and focus on the field of Economic History. His absolutely fantastic class and teaching methods, and precise and helpful encouragement in my research, reminded me how much I enjoyed reading History and showed me that I am capable of studying it rigorously with my quantitative background. If I have spent so many hours writing endless lines of code or waiting impatiently for boxes of archives, it has nonetheless been exciting because I am studying questions that I find fascinating. Although research can be frustrating at times, I still get a kick out of reading accounts of missionary travels or finding useful numbers in the archives. I can thus never thank enough the person who inspired me to go on this path.

I also wish to thank the other members of my jury whom I have not mentioned yet, Sascha O. Becker, Sergei Guriev, and Paul Seabright. I admire their research and I feel immensely honored that they took the time to read and comment my work, as well as to be present for the defense. Similarly, numerous researchers have helped me improve my articles during these past four years: Brian A'Hearn, Chris Bickerton, Ruben Durante, Quoc-Anh Do, James Fenske, Roberto Galbiati, Sergei Guriev, Pauline Grosjean, Élise Huillery, Thierry Mayer, Gerard Padro-i-Miquel, Thomas Piketty, Shanker Satyanath, Marc Sangnier, and Joachim Voth. Guillaume Laval et Étienne Patin m'ont introduit de manière express à la génétique des populations, un domaine avec lequel j'espère interagir en profondeur dans les années à suivre. There is also a number of younger economic historians with whom I have exchanged stimulating research ideas that have fed my own research, at either conferences, seminars, or "sometimes" cafés and bars : Cornelius Christian, Felipe Valencia-Caicedo, Jean Louis Keene, Johannes Bugge, Lara Tobin, Marc Goni, Leander Heldring, and -of course- Seyhun Orcan Sakalli.

I owe the people who welcomed me in Cambridge and in the LSE because they opened the possibility for new research and professional opportunities. Gerard Padro-i-Miquel and Jeremiah Dittmar helped me polish my first article at the LSE. Chris Bickerton helped me understanding the Oxbridge world, a help without which I would not be at Oxford now. I also wish to thank the students at POLIS, Weeda, Raphael, and Alexandra, who so kindly invited me to their colleges and satisfied patiently my curiosity for their Political-Science research.

Je ne suis pas très patiente, et parfois, j’aimerais tellement que l’inventée “phobie administrative” existe vraiment. Grâce à Alain Besoin, Cathy, Cécile, Leïla et Pilar, les habitants du département d’économie à Sciences Po vivons dans un petit îlot d’efficacité. Il est parfois difficile de se rendre compte que les procédures marchent bien quand ce qu’on souhaite c’est de les voir disparaître. Mais il ne faut pas oublier que c’est le résultat de leur travail qui nous épargne tellement de temps.

J’ai été heureuse chaque jour que j’ai gravi les quatre, puis cinq, étages de la rue des Saints Pères grâce à l’extraordinaire équipe de doctorants avec qui j’y ai partagé l’open-space. Deux jours après avoir commencé ma thèse, Elisa et moi avons partagé dans le rire nos talents d’ingénieur en essayant de comprendre comment marchaient nos nouvelles chaises de bureau. J’ai gardé la même chaise, avec les mêmes adaptations pour mon dos un peu trop fatigué pour mon âge, pendant toute la thèse. J’ai aussi entrepris des projets de collaboration extraordinaires avec Jean Louis, qui m’a montré la complexe vision de l’Union Européenne de Taylor Swift. J’attends avec impatience le début de nos recherches. Heureusement, aussi, que la thèse et les journées rue des Saints Pères ont été suffisamment longues pour que la timidité de Pierre D. ne m’ait pas empêché de faire la connaissance de quelqu’un dont la réserve n’ôte pas un millimètre d’intelligence ou d’humour ! Malgré mes origines et mes fréquentations de multiples sociétés étudiantes sud-américaines, je n’ai jamais rencontré quelqu’un avec une connaissance aussi exhaustive et précise du foot et du cyclisme. Je remercie aussi Jamil d’avoir été un excellent premier voisin et de m’avoir déculpabilisée de mon faible pour le power-nap, Eve avec qui nous avons pu partager nos visions de l’enseignement, ainsi que notre amour pour la cuisine, Clara W. pour nos discussions sur les théories du genre. Guillaume, bien sûr, mais que maintenant je remercie en dernier.

Bref, sans les moments de pause au déjeuner, à 17h, après les séminaires, ou parfois au O’Neil, écrire cette thèse aurait eu (beaucoup) plus de conséquences sur ma santé mentale. Merci donc à Assia avec qui j’aurai aimé pouvoir discuter davantage, Arthur, Cécile M., Charles, Clément, Dilan, Dylan, Eji, Emanuele, Étienne, François, Guillaume C., Guillaume V., Iván, Johannes, Léa, Marion, Meradj, Paul, Riccardo, Odran, Pierre G., Pierre C., et Xavier.

En dehors des départements, de l’open-space, des archives, et des bibliothèques, ma famille

et mes amis m'ont soutenu dans ce parcours sinueux de différentes manières. Mes parents, qui je ne sais trop comment ont réussi à inculquer quelques bonnes choses en moi. Je leur dois aussi ces magnifiques vices, l'esprit de contradiction et l'envie d'avoir raison, qui sont si souvent pratiques dans la recherche. Ils m'ont tous les deux fait découvrir des millions d'horizons, dont la cuisine, les langages et la fiction, vers lesquels je me tourne sans cesse pour me détendre. Lucia, dumela-ma, en qui je pense à chaque fois que je suis tentée d'abandonner quelque chose ou que je mange des tomates. Maxime, que j'ai décidé d'intégrer dans ma famille après 30 secondes de l'avoir rencontré. Je ne sais que dire non plus pour remercier Fabienne, Madelon, Titon, et Yohann -dont je suis tellement fière- qui ont une gentillesse sans limites, un sens de l'humour extraordinaire, et une joie de vivre si contagieuse. Rien n'aurait été possible sans eux.

Son demasiadas personas a las que tengo que darles las gracias! Pero no quiero olvidar a mis tíos, sobre todo Gungun y Nancy, que estuvieron tan pendientes de Lucía y de mí. Y aunque parezca sorprendente, debo agradecer a mi tía Beatriz, pues fue pensando en su labor social en Cali y otras ciudades de Colombia que empecé mi investigación sobre las misiones.

Agathe, Amélie, Anne-Laure, Clara, Cynthia, Ellen, Isabelle, Lara, Lucille, Pauline, Sarah, Scarlett, et Sohyun, que j'ai rencontrées à différents moments constituent une formidable équipe de soutien féminin. Chacune d'entre elles m'inspire car elles mènent à bout des projets ambitieux et fascinants dans des univers souvent trop masculins. En particulier Clara, avec qui nous avons partagé nos doutes et nos blagues douteuses depuis presque le début de nos études. Je remercie aussi Olivier, le seul excentrique de Bercy et plus converti au féminisme, qui est aussi la première personne à m'avoir parlé des enjeux de l'histoire de l'Afrique coloniale, ainsi que François avec qui j'ai partagé mon premier bureau et qui m'a vu lire des cartes à la loupe et faire des siestes entre les pages. Je dois aussi remercier Yannick, car nous avons grandi ensemble et qu'il m'a appris à coder, ainsi qu'Élie car nous sommes devenus économistes ensemble. Dans la même veine, toute l'équipe du C-5 m'a décomplexée d'une bonne fois pour toutes de ma geekitude.

Enfin, je tiens à remercier Guillaume qui réussi à me faire rire sous toutes conditions, et qui en plus apprécie mon caractère révolté ! C'est à tour de rôle que la recherche nous rend passagèrement fous ou pessimistes, mais en nous rassurant et en profitant de la vie ensemble,

nous avançons. Il faut bien que nous remercions la thèse et Sciences Po, puisque c'est là que l'on s'est croisés.

Santé.



## Résumé

Les comportements politiques et sociaux tels que la participation politique, la confiance en autrui, l'engagement collectif, la prévention en santé, ou les attitudes vis à vis de la contraception, peuvent persister pendant de très nombreuses années. Cette thèse présente des travaux qui explorent et quantifient rigoureusement des instances de persistance dans ces comportements, en utilisant de nouvelles sources de données historiques et contemporaines. Les travaux présentés dans cette thèse contribuent à la littérature de trois manières différentes. En premier lieu, ils présentent une nouvelle base de données sur la présence des missionnaires chrétiens en Afrique et les investissements qu'ils y ont conduit. Cette base de données est unique en ce qu'elle est entièrement géocodée et présente des données à un niveau de désagrégation très fin. En deuxième lieu, ces travaux mettent en avant de manière originale des canaux de persistance dans le développement qui ne sont pas attribuables aux différences institutionnelles. En troisième lieu, en analysant la marge intensive de la diversité, ces travaux proposent aussi une nouvelle manière d'aborder la question de l'endogénéité dans l'étude du rôle économique de la diversité des origines dans une société.

La première partie de cette thèse analyse l'effet persistant sur le développement de l'activité missionnaire en Afrique subsaharienne. Je présente ici la recherche que j'ai menée avec Julia Cagé. Cette recherche utilise notamment notre nouvelle base de données compilée et géocodée à partir de l'Atlas des missions protestantes publié en 1903 par le *Student Volunteer Movement for Foreign Missions, New York* (Beach, 1903). Cette partie se décompose en deux chapitres.

Le premier chapitre présente les effets de long-terme de l'introduction de l'imprimerie au XIX<sup>ème</sup> sur le capital social aujourd'hui. Les missionnaires protestants furent les premiers à importer l'imprimerie en Afrique subsaharienne et à y garantir l'accès aux populations locales. Aujourd'hui, parmi toutes les régions qui ont connu la présence de missionnaires chrétiens, celles proches de l'imprimerie en particulier ont davantage de consommation de média et un capital social plus important. La proximité à d'autres investissements n'a pas un effet comparable.

Le deuxième chapitre présente les effets de long terme des investissements des missionnaires sur la prévalence du VIH/SIDA. Avec des données biologiques contemporaines, nous

montrons que la proximité à une mission est associée à une prévalence plus forte, conditionnellement aux facteurs de risque usuels (urbanisation, ou densité routière). Cependant, être proche d'un investissement en santé fait par les missionnaires, comme un hôpital ou un dispensaire, est associé à une prévalence plus faible. Il y a donc un effet conflictuel du rôle des missions sur la prévalence du VIH/SIDA aujourd'hui. D'une part, les investissements en santé réduisent la prévalence aujourd'hui; d'autre part, la présence missionnaire a aussi influencé les comportements de prévention en diminuant l'acceptabilité des contraceptifs et de l'utilisation de préservatifs.

La deuxième partie de cette thèse étudie les conditions sociales qui font que les différences d'origine peuvent devenir une barrière à la réussite économique aux États-Unis. Les États-Unis sont un choix naturel dans l'étude de la diversité des origines, parce que c'est un pays qui a été peuplé par des vagues de migrations diverses et successives pendant une période relativement courte. En outre, le Bureau du recensement américain a gardé la trace des origines de ses habitants depuis 1850.

J'utilise les distances génétiques bilatérales entre les origines déclarées comme mesure de la marge intensive des différences d'origines. Ces distances reflètent l'étendue des échanges de populations dans le passé entre les groupes concernés. Le troisième chapitre explique en détail la construction de ces mesures génétiques, ainsi que la mesure dans laquelle elles reflètent des différences culturelles entre les groupes.

Le quatrième chapitre présente les résultats de ma recherche sur l'effet sur le succès économique d'une individu d'être d'une origine différente par rapport à la moyenne locale. Cette "différence d'origines" est mesurée par les distances génétiques mentionnées ci-dessus. Je montre que ces différences n'ont pas d'effet systématique sur le succès économique. En revanche, l'effet est poussé à la hausse (vers un effet positif), lorsque les sociétés sont plus probablement enclines à avoir confiance en autrui. Au contraire, lorsque les sociétés sont plus probablement enclines à la xénophobie ou à être fermées à l'altérité, l'effet est poussé à la baisse (vers un effet négatif).

# Abstract

Individual and social behaviors, such as voting patterns, the ability to trust others, participation in the community, health-seeking behaviors, or attitudes towards contraception can persist over many years. Relying on new historical and contemporary data sources, the works compiled in this thesis aim at explaining and quantifying rigorously instances of persistence in behaviors and preferences. This work presents three contributions to the literature. Firstly, it compiles a new micro-level geocoded historical datasource on missionary activity in sub-Saharan Africa. Secondly, it has elucidated, in original ways, channels of persistence in development, which are not attributable to formal institutions. Thirdly, it has analyzed the intensive margin of diversity, providing a new way to address the endogeneity issue in the study of the role of diversity on income.

The first part of the thesis is a work on the long-lasting influence of missionary activity in sub-Saharan Africa. I present there the research I have conducted with Julia Cagé. It relies a new database that we constructed and geocoded based on the *Geography and Atlas of Protestant missions* (Beach, 1903). This research is developed in the two first chapters of the thesis.

Chapter 1 explains the long-term consequences of the introduction of the printing press in the nineteenth century on social capital nowadays. In sub-Saharan Africa, Protestant missionaries were the first both to import the printing press technology and to allow the indigenous population to use it. Within regions located close to missions, we show that proximity to a printing press is associated with higher media consumption and social capital. Proximity to other missionary investments does not exhibit such consistent effects on contemporary social capital outcomes.

Chapter 2 investigates the long-lasting consequences of missionary investments on HIV prevalence today. Using contemporary individual-level biological data, we show that proximity to historical mission settlements is correlated with high HIV prevalence today, conditional on the already identified risk factors for prevalence such as urbanization and road density. However, proximity to a health investment is associated with lower prevalence rates. There is thus a conflicting effect of missionary activity on HIV prevalence today. On the one hand, health-related investments still tend to decrease prevalence today; but on the other hand,



missionary presence also influenced perceptions on condom use, female contraception, and the stigma on HIV which in turn tends to be associated with higher prevalence today.

The second part of this thesis studies the social conditions under which ancestry differences act as a persistent barrier for individual earnings in the United States. The United States is a natural choice to study the persistent effect of ancestry on income, as this is a country that was populated by subsequent waves of migration from different origins during a relatively short period of time. Moreover, the American Census Bureau has recorded data about ancestries since the 1850s.

I use the pairwise genetic distance between ancestries to capture the intensive margin variation in ancestry differences. These distances capture the extent of population exchanges that each pair of ancestries has had in the past. Chapter 3 explains with more details the construction of the genetic distance, as well as why and to which extent it captures cultural differences between groups.

Chapter 4 presents the results of my research on the role of distance in ancestry at the local level as a determinant of earnings. I show that individuals' distance to the ancestries present in her area has heterogeneous predictive power. There is no baseline robust correlation between distance to other ancestries and individual economic success. However, for all the measures used, living in an area where other individuals come from trusting ancestries pushes the effect upwards. On the contrary, living in an area where other individuals come from xenophobic ancestries pushes the effect downwards. First-generation migrants seem to be less likely to success the more culturally distant they are, but this effect vanishes quickly as time spent in the USA increases.





# Contents

<b>Remerciements</b>	<b>5</b>
<b>Résumé</b>	<b>11</b>
<b>Abstract</b>	<b>13</b>
<b>Contents</b>	<b>17</b>
<b>Introduction</b>	<b>25</b>
<b>I The long-term consequences of missionary investments in sub-Saharan Africa</b>	<b>37</b>
<b>1 The long-term consequences of the printing press</b>	<b>39</b>
1.1 Introduction . . . . .	39
1.2 Historical background and persistence . . . . .	43
1.3 Data and missionary locations . . . . .	48
1.3.1 Data . . . . .	48
1.3.2 Determinants of mission location and investments . . . . .	52
1.4 Empirical analysis . . . . .	56
1.4.1 Specification and identification strategy . . . . .	56
1.4.2 OLS Estimation . . . . .	59
1.4.3 Endogenous selection of missions into printing and other investments	64
1.4.4 Robustness checks . . . . .	70
1.4.5 Additional results . . . . .	71

1.5	Discussion . . . . .	74
1.6	Conclusion . . . . .	76
<b>2</b>	<b>Sex and the Mission</b>	
	<b>The conflicting effects of early Christian investments on sub-Saharan Africa's HIV epidemic</b>	<b>79</b>
2.1	Introduction . . . . .	81
2.2	Historical background . . . . .	85
2.2.1	Missions and early health investments . . . . .	85
2.2.2	The Christian influence and the focus on abstinence . . . . .	86
2.3	Data Sources and Description . . . . .	88
2.3.1	Contemporary Data . . . . .	88
2.3.2	Historical Data . . . . .	90
2.3.3	Determinants of mission location and investments . . . . .	91
2.4	Empirical Analysis . . . . .	94
2.4.1	Specification . . . . .	94
2.4.2	Results . . . . .	95
2.5	Robustness checks and Discussion . . . . .	100
2.6	Conclusion . . . . .	104
<b>II</b>	<b>The role of ancestry to explain individual economic success at the local level in the United States</b>	<b>111</b>
<b>3</b>	<b>What Population Genetics can and cannot tell us about societies</b>	<b>113</b>
3.1	Within-population measures: using genetic diversity in the Social Sciences . . . . .	114
3.1.1	Heterozygosity . . . . .	115
3.1.2	Heterozygosity and Genetic Drift . . . . .	115
3.1.3	Heterozygosity in the Social Sciences . . . . .	117
3.2	Between Population Measures: is genetic distance a proxy for cultural distance?121	
3.2.1	Gene flow . . . . .	121
3.2.2	Genetic distance . . . . .	122

3.2.3	Genetic and cultural distances . . . . .	125
3.3	Conclusion . . . . .	128
<b>4</b>	<b>Achieving the American Dream:</b>	
	<b>Ancestry distance, cultural diversity, and economic success</b>	<b>133</b>
4.1	Data . . . . .	140
4.1.1	American Community Survey . . . . .	140
4.1.2	Genetic Distance as a Proxy for Cultural Distance . . . . .	140
4.1.3	Genetic $F_{ST}$ and Cultural Distance . . . . .	142
4.1.4	Diversity . . . . .	145
4.1.5	Summary Statistics . . . . .	146
4.2	Cultural Distance and Economic Success . . . . .	146
4.2.1	Specification . . . . .	146
4.2.2	Baseline Results . . . . .	149
4.2.3	Endogenous Location of Migrants . . . . .	150
4.3	Channels . . . . .	152
4.3.1	First and Older Generation Migrants . . . . .	152
4.3.2	The Role of Attitudes to Explain the Effect of Cultural Distance on Income . . . . .	158
4.4	Robustness . . . . .	164
4.4.1	Alternative Outcomes . . . . .	164
4.4.2	Alternative Measures of Cultural Distance . . . . .	168
4.4.3	Polarization . . . . .	168
4.5	Conclusion . . . . .	169
<b>A</b>	<b>Appendix to Part 1</b>	<b>171</b>
A.1	Data . . . . .	171
A.1.1	Data on Mission Centers Location and Investments . . . . .	171
A.1.2	Afrobarometer Data . . . . .	174
A.1.3	Controls . . . . .	174
A.1.4	Newspaper Market . . . . .	176

A.1.5	Descriptive Statistics . . . . .	182
A.2	Matching: Additional tables . . . . .	187
A.3	Using selection on observables to assess the bias from unobservables: Technical details and results . . . . .	187
A.4	Additional Results . . . . .	194
A.5	Robustness Checks . . . . .	198
A.5.1	Distance . . . . .	198
A.5.2	Additional Robustness Checks . . . . .	200
<b>B</b>	<b>Appendix to Part 2</b>	<b>207</b>
B.1	Data . . . . .	207
B.1.1	American Community Survey . . . . .	207
B.1.2	Cultural Distance . . . . .	209
B.2	Additional Results . . . . .	211
B.2.1	Selection on Unobservables . . . . .	211
B.2.2	Alternative Measures of Cultural Distance . . . . .	213
B.2.3	Polarization . . . . .	219
	<b>List of Figures</b>	<b>221</b>
	<b>List of Tables</b>	<b>223</b>
	<b>Bibliography</b>	<b>227</b>







“Even a very rich person who is prevented from speaking freely, or from participating in public debates and decisions, is *deprived* of something that she has reason to value. The process of development, when judged by the enhancement of human freedom, has to include the removal of this person’s deprivation. Even if she had no immediate interest in exercising the freedom to speak or participate, it would still be a deprivation of her freedoms if she were to be left with no choice on these matters. Development seen as enhancement of freedoms or civil rights cannot address such deprivations. The relevance of the deprivation of basic political freedoms or civil rights, for an adequate understanding of development, does not have to be established through their indirect contribution to *other* features of development (such as the growth of GNP or the promotion of industrialization). These freedoms are part and parcel of enriching the process of development.”

*Development as Freedom*, Sen (1999, p.37)



# Introduction

Individual and social behaviors, such as voting patterns, the ability to trust others, participation in the community, health-seeking behaviors, or attitudes towards contraception can persist over many years because external factors such as institutions or geography will constraint choices over generations (Engerman and Sokoloff, 1997; Acemoglu et al., 2001), or because tastes an human capital, that determine those behaviors, are transmitted by parents (Bisin and Verdier, 2001; Glaeser et al., 2004; Algan and Cahuc, 2010).

Relying on new historical and contemporary data sources, the works compiled in this thesis aim at explaining and quantifying rigourously instances of persistence in behaviors and preferences. I identify persistent patterns in political and economic behaviors, such as media consumption, trust, or health seeking behaviors. The first part of the thesis delves into the understanding of the long-lasting influence of missionary activity in sub-Saharan Africa. This research shows how historical investments made by missionaries persistently modified media consumption and health-seeking behaviors. The second part studies the social conditions under which ancestry differences act as a barrier for individual earnings in the United States.

Because there is no perfectly measurable definition of development, and my work is mostly an empirical effort of quantification, I will avoid using the term as much as possible in my chapters. However, when outlining my work and its relation to the literature in this introduction, I was compelled to include the outcomes I study as part of the concept of development. The outcomes considered in this dissertation are diverse, as I measure media consumption, trust, political participation, engagement in the local community, HIV prevalence, attitudes towards contraception, and earnings. The choice of outcomes was usually driven by the specific nature of the questions treated; within the possible range of outcomes, data availability

constrained my choice. Relying on Amartya Sen’s approach, I consider “development” as the process of “expanding the real freedoms that people enjoy” (Sen, 1999, p. 3). The broad range of outcomes I consider constitute a subset of means that are relevant in view of increasing human freedoms. Individuals living places with similar levels of income can enjoy different levels of freedom depending on the political rights they can enjoy, their life expectancy, or the distribution of opportunities. Individuals can express their political preferences in the media, and without fear if they trust individuals in their community. Similarly, health seeking behaviors proxy for the capacity of individuals to care for their own lives. These outcomes do not translate into development *per se* in the view of Amartya Sen, because they are merely means to achieve it. However, as means of development, understanding their historical determinants in particular contexts is indeed a question of relevance for the field of Development Economics.

The study of the persistent economic effect of historical events has been closely linked to the literature on the institutional roots of development. In his lecture notes on Political Economy, Acemoglu highlights two sources of persistence in institutions. Firstly, political institutions are durable because a “necessarily large” shift in the distribution of political power is necessary to change them. Secondly, the wealth of a particular group will de facto increase its political power, which will in turn reinforce its claim over political and economic institutions in its favor (Acemoglu, 2011, p. 59). This idea is also developed in *Why Nations Fail* (Acemoglu and Robinson, 2012, p. 407), as the authors state:

“As virtuous cycles make inclusive institutions persist, vicious cycles create powerful forces towards the persistence of extractive institutions. History is not destiny, and vicious cycles are not unbreakable (...). But they are resilient. They create a powerful process of negative feedback, with extractive political institutions forging extractive economic institutions, which in turn create the basis for the persistence of extractive political institutions.”

Therefore, institutions have persistent effects over time because they can set societies in particular stable equilibria. The stability is due to the feedback loop generated between political and economic power.

Institutions are not the only factor in a society that can have persistent consequences for development. Historical investments and trade can also have durable consequences if they change preferences or market access, which in turn creates a feedback of increasing

## Introduction

demand for what the initial investment supplies (Feyrer and Sacerdote, 2009; Huillery, 2009). Migration might also have substantial consequences, notably when it is high enough to lead to a significant change in the levels of human and social capital (Glaeser et al., 2004, among many others).

The empirical literature on the long-term determinants of development faces the difficulty of establishing convincing empirical evidence because the data typically vary at the country-level. This data constraint increases in turn the challenge of disentangling the effect of historical events from other geographical or economic confounding factor. A recent literature aims at surpassing this identification challenge by focusing on original data that exhibits variation in the subnational level to identify long-lasting consequences of institutions (Nunn, 2008; Dell, 2010; Nunn and Wantchekon, 2011; Michalopoulos and Papaioannou, 2013, 2014), or colonial investments (Huillery, 2009).

Inspired by the micro-econometric approach of the aforementioned literature, my research has focused (although not exclusively either!) on the historical determinants of behaviors such as media consumption, health-seeking behaviors, or trust. The thesis is divided in two main parts corresponding to the two main cases I study. The first one explores the long-term effects of missionary investments in sub-Saharan Africa. The second one analyzes is the explanatory power of individual ancestry on individual economic success in the USA.

In the first part of the thesis, I present the research I have conducted with Julia Cagé. We analyze subnational variation in the provision of Protestant missionary investments in the early twentieth century. Missionary activity supplied extensive health, education, and cultural services in sub-Saharan Africa, without changing the formal institutions decided by the colonial rule. Their presence varied due to resource and geographical constraints. More importantly, there was also heterogeneity among missions in the type of investments they provided. We exploit this “within-mission” variation in investments to identify the different effects of investments on the printing press, schooling, and health on media consumption, social capital, and health seeking behaviors.

I also exploit variation at the sub-national level in the second part of the thesis, when analyzing the effect of ancestry differences on economic success in the United States. Migrants bring with them culture, and human and social capital, which are persistently transmitted

across generations (Glaeser et al., 2004; Algan and Cahuc, 2010). For this reason, an institutionally unified population as diverse in origins as the North American one, is particularly relevant to understand the persistent effect of diversity on economic success. Using the sub-national variation in ancestry recorded in the American Community Survey, I show that ancestry is a robust predictor of economic success among first generation migrants, but the effect vanishes quickly as time spent in the USA increases. The effect also depends on the communities' level of social capital.

My research provides evidence for the long-lasting consequences of infrastructure and population movements (migration, but also missionary settlements) taking formal institutions as given. Although I focus the persistence of these historical events, my approach is different from the standard institutional approach inasmuch as it does not seek to find institutional variation, but variation in historical investments or culture that durably affected development. In that sense, this work is more closely linked to the literature on the consequences of human capital on development (Glaeser et al., 2004, and subsequent work) than to the institutional approach to development (Acemoglu et al., 2001, and subsequent work).

However, I do not claim that the latter approach is a unique (nor better) explanation of the persistence observed in development patterns than the former. I have proven that people's cultures, beliefs, and preferences are likely to persist over time but I have not provided any proof that institutions do not. Moreover, as mentioned before, the recent literature has also made a strong case for the idea that formal institutions can persistently change development paths at the local level, in particular when the State's capacity is sufficiently strong to apply the rule of law in all its territory (Michalopoulos and Papaioannou, 2014). On the contrary, I would rather claim that since economic interactions are embedded in social relations, then historical events that affect people's preferences, culture, or constraints on behaviors (like institutions) can have long-lasting effects on development.

I will now present a more precise description of each chapter's content as well as their contributions to the literature.

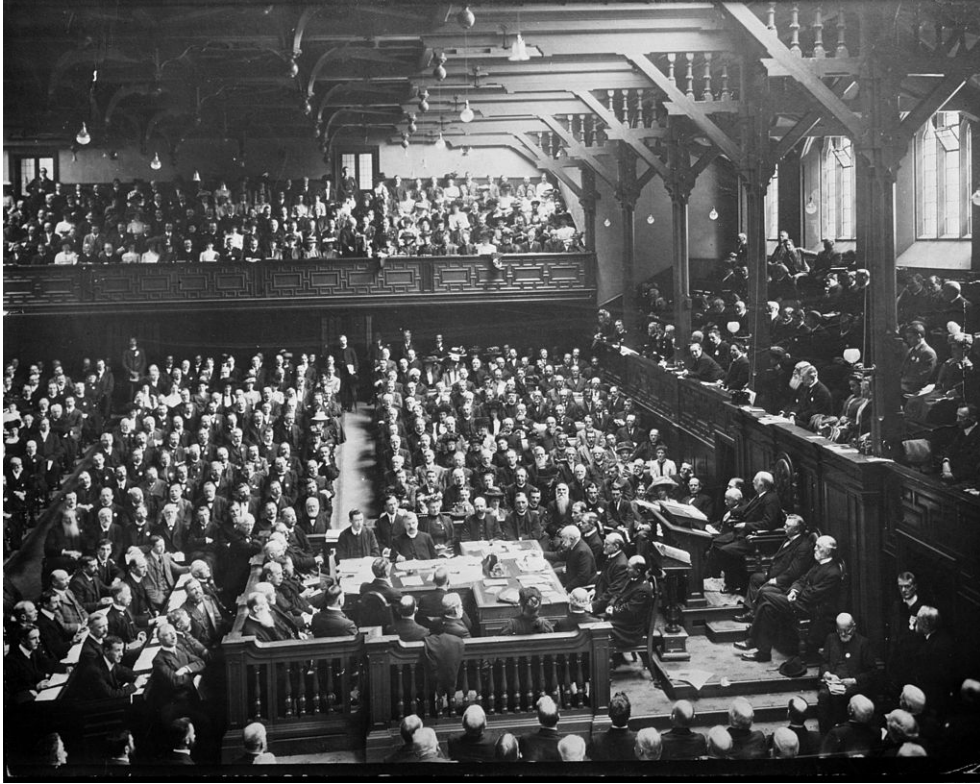
## **The long-term consequences of missionary activity in sub-Saharan Africa**

The first part of my thesis is the result of my now five-years long collaboration with Julia Cagé. This research investigates the long-term consequences of the investments undertaken by Christian missionaries in sub-Saharan Africa in the early twentieth century.

The role of colonization on the persistence of development has been widely studied in Economic History and Development. However, colonial administrations were not the only ones investing and influencing the local social and cultural organizations. In South-America, for instance, the oldest and largest schools were not built by the Spanish Crown, but by Jesuit missionaries. Prominent figures in South-American History were educated in Bogota's "Colegio Mayor de San Bartolomé", the Jesuit school built in 1604, which was one of the first in Latin-America. Antonio Nariño was educated there, he later translated from French to Spanish the "Declaration of the rights of Man and of the Citizen", which he printed and distributed as tracts. He later became one of the Leaders of the independence movement of the Nueva Granada. Valencia-Caicedo (2014) recently shed light on the persistent effects of Jesuit investments in education amongst the Guaraní people in modern-day Argentina, Brazil, and Paraguay. Not only do they tend to have higher income, but they also exhibit higher non-cognitive ability and collaborative behavior.

The second wave of European colonization in sub-Saharan Africa led to a similar pattern as missionary activity, and this time also Protestant missionary activity, exploded in the late nineteenth and early twentieth Century in sub-Saharan Africa, India, and East Asia. In 1903, there were more than 4000 mission stations from 257 societies operating in the world. Among them, 723 were located in sub-Saharan Africa. Missionary activity also benefited from improved means of transportation and communication, contributing to an unprecedented increase in their activities around the world. The amount of resources devoted to organizing Edinburgh's World Missionary Conference in 1910, is further proof of the commitment and effort devoted to Christian missionary work. More than 1215 representatives from Europe and North-America met in Edinburgh to discuss about the challenges faced by missionary activity worldwide (Stanley, 2009). Figure 1 shows a picture of one of the meetings during this conference, and illustrates the large number of people involved.





**Figure 1** – The World Missionary Conference in Edinburgh, 1910

**Note:** This picture was taken during the World Missionary Conference in Edinburgh, 1910. This conference aimed at discussing the challenges and goals that missionary activity faced and increasing missionary cooperation in the field.

Comparing regions depending on whether they were affected by missionary activity or not is highly illustrative of the effect of Christianity -and in particular protestantism- on economic development, a relationship that was early studied by Weber (1930), and that Woodberry (2004, 2012) has more recently analyzed for regions where missions settled. Most missionaries invested in education and literacy, because the Protestant principle of the “Sola Scriptura” requires all Christians to read the Bible. This specificity of Protestantism has been studied in the recent Economic History literature as an explanatory factor of urbanization that worked through the increase in human capital in Germany (Becker and Woessmann, 2009; Cantoni, 2015), Guatemala (McCleary and Pesina, 2012), and China (Bai and Kung, 2011). This literature is mostly based on the comparison between regions that converted to Protestantism and those that did not.

Our research is different inasmuch as it focuses only on regions that converted to Protestantism because missionary activity was also highly heterogeneous. Missions invested in

## Introduction

territories that were occupied by the colonial powers, but they benefited from a certain degree of independence in the choice and management of their investments. Depending on their means and the type of skills of their missionaries, missions would specialize on different activities. In sub-Saharan Africa, the most common ones were health, education or printing. The work Cagé and I have conducted has aimed at exploiting this heterogeneity to assess precisely the effects of these investments on different dimensions of development such as media consumption, social capital, and health.

Our research uses a new database that we constructed based on the *Geography and Atlas of Protestant missions* (Beach, 1903). This atlas contains maps locating all the mission stations around the world in 1903 and inventories in detail all the investments conducted at each station. We geocoded all this information to obtain a dataset including, for each mission station, the geographic location and its characteristics, as well as the printing-, educational- and health-related investments. Figures A.1 and A.2 in the Appendix provide examples of the maps and indexes from which the dataset was created.

Chapter 1 is the result of our research on the long-term consequences of the introduction of the printing press in the nineteenth century on civic and social capital nowadays. The research paper is to be published in July 2016 (Cagé and Rueda, 2016). In sub-Saharan Africa, Protestant missionaries were the first both to import the printing press technology and to allow the indigenous population to use it. These missionaries pioneered in the development of a written tradition for sub-Saharan African languages; they formalized indigenous languages and printed Bibles and educational material in these languages. Protestant missionaries were thus among the first to bring the printing press to sub-Saharan Africa; on the contrary, Catholic missionaries had no role in bringing the printing press. This chapter shows that, within regions located close to missions, proximity to a printing press is associated with higher media consumption and social capital. Proximity to other missionary investments does not exhibit such consistent effects on contemporary social capital outcomes. Our results are robust to a variety of identification strategies.

Chapter 2 investigates the long-lasting consequences of missionary investments on HIV prevalence today in sub-Saharan Africa. The history of modern medicine in sub-Saharan Africa is intrinsically linked to the development of missionary activity. Missionaries were

the first and only ones to invest in modern medicine there until the middle of the twentieth century. However, we shall not forget that the first objective of missionary activity was to evangelize and that most of their activities were conducted in view of achieving this goal. Health services were part of the missionary work because the notion of healing is so deeply rooted in the Christian dogma. As an example of the overlap between the notion of conversion and the one of healing, we can Gospel of Matthews 10:1 that reads: “Jesus called his twelve disciples to him and gave them authority to drive out impure spirits and to heal every disease and sickness”.

The tight links between subjective beliefs and morality on the one hand, and medical care on the other hand, was problematic for the efficient treatment of sexually transmitted infections in sub-Saharan Africa since the colonial times. In Uganda, there was a large increase in the prevalence of syphilis from the mid nineteenth century, which led to one of the first public health campaigns launched in sub-Saharan Africa by the colonial powers (Setel et al., 1999, p.99). The priority of the campaign, that was agreed between European missionaries, colonial administrators, and the African chiefs, was to fight the innate sinfulness of the African society Vaughan (1991) to eradicate the *endwadde ez’obukaba*, the “disease of immorality”.

The public campaign against syphilis in Uganda is illustrative of the tension between medical work and the Christian morality in the early twentieth century. The second chapter of this thesis aims at quantifying whether this tension is also relevant in the fight against HIV/AIDS. Duflo et al. (2015) show that the public health campaigns against HIV/AIDS in Kenya (and in most sub-Saharan African countries, they claim) are still largely conducted by Christian associations and focus on abstinence and monogamy. Using contemporary individual-level biological data, we show that proximity to historical mission settlements is correlated with high HIV prevalence today, conditional on the already identified risk factors for prevalence such as urbanization and road density. However, among regions historically close to missionary settlements, proximity to a health investment is associated with lower prevalence rates. There is thus a conflicting effect of missionary activity on HIV prevalence today. On the one hand, health-related investments still tend to decrease prevalence today; but on the other hand, missionary presence also influenced perceptions on condom use,

female contraception, and the stigma on HIV which in turn tends to be associated with higher prevalence today.

### **Ancestry and Individual Economic Success in the United States**

My research on the long-term consequences of the printing press in sub-Saharan Africa introduced me to the theories of Anderson (1991), who claims that the printing press is also an explanatory factor of nationalism, as a catalyst of the consolidation of “imagined communities” such as the nation or the ethnicity. What is the role of these imagined communities on economic development? Do they increase development by fostering the accumulation of social capital? Or on the contrary, does the strong identification to the nation or the ethnicity increase the probability of conflict by increasing fractionalization? The quest for answers to these broad questions took me on a new path of research, separate from the analysis of missionary activity. The second part of the thesis presents the results of this second line of research.

The particular question that I found necessary to tackle is the one of the impact of diversity in those imagined communities -ethnicities, nations, or ancestries- as fundamental determinants of individual economic success. This is a matter of crucial economic, political, and historical relevance as, on the one hand, ethnic and ancestry diversity sharply increased in the twentieth century (Putnam, 2007); and on the other hand, the twentieth century’s extreme-right wing movements always built their political discourse upon the fear of such increasing diversity. Is it true that people from different ancestries or culture are less economically successful? What is the persistence of any effect of diversity on development? Does the effect vanish as individuals are integrated in the community or can it last for centuries (Ashraf and Galor, 2013)?

Analyzing the role of diversity on economic performance is a task that economists have conducted several times over the years, and which faces a large number of challenges. The first one, is the problem of causality. Does diversity cause growth or do prosperous places attract people from a larger number of ancestries? May it also be that diversity increases the probability of conflict as shown by the work of Alesina et al. (2003); Montalvo and Reynal-Querol (2005); Esteban et al. (2012)? The second challenge, is the problem of measurement,

how shall we define diversity? Should we focus on ethnicity, ancestry, religion, language, or a combination of all those things? The recent literature has shown that these groups do not necessarily overlap (Desmet et al., 2015).

Given the number of challenges faced, it seemed necessary to simplify the question, which I did by restricting the analysis to a single country. The United States was a natural choice, as this is a country that was populated by subsequent waves of migration from different origins during a relatively short period of time. Actually, according to the Encyclopedia Britannica, the Great Atlantic Migration from Europe to the USA and from 1840 to the Great War, is the largest migration movement in history. Moreover, the American Census Bureau has recorded data about ancestries since the 1850s. Therefore, using the words of Fulford et al. (2015), the United States are a unique laboratory for any analysis on the role of diversity in ancestry in a given society. My research thus investigates the role of diversity in ancestry as a determinant of individual economic success.

To overcome the problem of reverse causality, I rely on measures of diversity that have an intensive margin variation. This means that in two places with the same number of different ancestries, there can be variation in the measure of diversity depending on how different the ancestries are. Following Spolaore and Wacziarg (2009), I use the pairwise genetic distance between ancestries to capture the intensive margin variation in ancestry differences. It is crucial to note that the genetic distance is based on neutral changes in DNA, which means sections of the DNA that do not influence physical characteristics. What it captures instead is the degree of genealogical relatedness between ancestries. The measure of genetic distance captures the extent of population exchanges that each pair of ancestries has had in the past. Chapter 3 explains with more details the construction of the genetic distance, as well as why and to which extent it captures cultural, linguistic, or religious distances between groups.

Chapter 4 presents the results of my research on the role of distance in ancestry at the local level as a determinant of earnings. I received help for this work from two population geneticist -Guillaume Laval and Etienne Patin- who guided my introduction to the notions of population genetics used in Economics. I show that individuals' distance to the ancestries present in her area has heterogeneous predictive power. There is no baseline robust correlation between distance to other ancestries and individual economic success. However, for all the

## Introduction

measures used, living in an area where other individuals come from trusting ancestries pushes the effect upwards. On the contrary, living in an area where other individuals come from xenophobic ancestries pushes the effect downwards. First-generation migrants seem to be less likely to succeed the more culturally distant they are, but this effect vanishes quickly as time spent in the USA increases.

In some regions, ancestry distance can be associated with higher levels of income, even when controlling for ancestry-specific fixed effects. This is more likely to be the case in regions populated with individuals from highly trusting ancestries. On the contrary, the correlation is more likely to be nil or negative in regions where individuals come from highly xenophobic ancestries. If we interpret distance in ancestry as cultural distance, then these results go in line with a non-static view of the role of culture as a barrier or catalyst of economic success. Economic integration is embedded in the society, and as such, it is the result of an interactive process. If cultural distance can become an economic advantage at the micro level, it will only be so in an environment that is open to diversity.

As opposed to the research on missionary activity activity, this part of the thesis does not focus on the long-term persistence of one particular historical event. On the contrary, it is conducted with modern data. It investigates the conditions under which diversity in ancestry -which can date for many years as declared the ancestry can date from several generations before- is still a relevant predictor of economic success at the individual level. I show that in an institutionally unified country, the relevance of ancestry as a predictor of economic success depends on the ability of people to trust others and be open to alterity. Negative attitudes towards alterity increase the relevance of ancestry as a negative predictor of income.



## Part I

# The long-term consequences of missionary investments in sub-Saharan Africa





# Chapter 1

## The long-term consequences of the printing press

### 1.1 Introduction

This article studies the long-term effects of early Protestant missionary activity in sub-Saharan Africa on civic and social capital nowadays. In particular, we investigate the long-term consequences of the early introduction of the printing press in the 19th century. Civic and social capital can have a profound effect on democracy and development.

In the 19th century sub-Saharan Africa, missions invested in numerous activities, among which education, health and printing. We built and geocoded an entirely new dataset of Protestant mission settlements. For each settlement, we document the exact geographic location, the educational, health-related and printing investments, as well as geographic and historical characteristics.<sup>1</sup> Producing these data is our first contribution.

Our second contribution is to identify the long-term effects of these investments. We investigate the effect of proximity to a range of mission characteristics on a range of contemporary outcomes. Using an original empirical strategy relying on our reading of history, we focus on the introduction of the printing press and its effects on contemporary newspaper readership, education, local civic engagement and social capital.

---

<sup>1</sup>Protestant missionary activity played a central role in the development of a written tradition in sub-Saharan Africa. Because they needed to print Bibles and educational material, Protestant missionaries were among the first to bring the printing press to Africa.

A wide literature suggests that higher social capital, and in particular trust, leads to higher economic activity (Guiso et al., 2008; Tabellini, 2010) as well as to higher well-being (see Algan and Cahuc, 2014, for a survey of this literature). Similarly, newspaper readership is a relevant proxy for social capital (Putnam, 2000; Guiso et al., 2004), and it is consistently associated with political participation (Gentzkow et al., 2011; Cagé, 2014) and voter knowledge (Snyder and Strömberg, 2010; Casey, 2015).

Using contemporary individual-level data from the Afrobarometer, we find that proximity to the closest location of a mission with a printing press has a positive and statistically significant impact on the probability of reading the news, our first proxy for social capital. A one-standard deviation increase in the proximity to a mission with a printing press increases the probability of reading the news on a monthly basis by around 5.3% of a standard deviation, trust by 8.1%, and education by 6.3%. Finally, we show that in democracies, proximity to a mission with a printing press increases political participation at the local level nowadays. In contrast, proximity to a mission without a printing press – whether Protestant or Catholic – has no significant impact on newspaper readership.

We then investigate the extent to which different types of missionary investments affect different contemporary outcomes. We find a positive correlation between proximity to a mission with a school and education nowadays. However, we find no persistent effect of proximity to a mission with a hospital on contemporary health-related outcomes.

Our econometric analysis attempts to move beyond two forms of selection. First, historical and geographical characteristics might have determined mission station location, preventing the comparison between regions close and far from these settlements. Protestant missionaries did choose to locate in geographically favored areas (Johnson, 1967; Nunn, 2010). Second, Protestant stations invested in different activities such as printing, health and education. There may be endogenous selection of missions into these activities.

To address selection from missions' location, we restrict our sample to regions *near* historical mission settlements. Because regions near Protestant missions shared similar geographic, institutional and cultural environments, this restriction isolates the specific effect of the investments from other possible long-term determinants of civic and social capital embedded in specific mission locations.

To address selection of missions into different investments, we first control for observable covariates. The set of observable covariates include geographic and historical characteristics as well as distance to historical mission stations that invested in health, educational facilities and printing presses. We then develop a matching strategy that aims at isolating the effect of proximity to a mission with a specific investment (printing press, school or hospital) from the effect of proximity to a mission with similar characteristics, but without this specific investment. The results are robust to this approach and of similar order of magnitude. Finally, we use insights from Altonji et al. (2005) and Oster (2013) to assess the bias due to unobservables using the sensitivity of the treatment to added controls. From this approach, it seems unlikely that the entire estimated effect of the distance to the printing press is driven by unobserved variables.

Despite our attempts to control for observable factors, our estimates might be driven by unobserved determinants of long-term development and proximity to a historical mission settlement endowed with a printing press. The long-term effects of printing missions on civic and social capital seem specific to missions with presses and is not related to other observed mission characteristics. Moreover, information on the number of publications printed at the mission press until 1923 suggest that the impact of the distance to a printing press on newspaper readership may be explained by the development of publishing activities. Although we cannot ultimately separate the effect of printing *per se* from other mission characteristics, a causal interpretation of the results seems plausible.

**Related literature** The long-term consequences of religious incentives for economic success have been widely studied in the social sciences, the most well-known theory being Max Weber’s “Protestant Ethic” (Weber, 1930). According to the principle of the *Sola Scriptura*, central to the Protestant doctrine, every Protestant should be able to read the Bible. Recent work has emphasized this incentive to increase literacy as an alternative explanation for the economic success of regions that first converted to Protestantism (Becker and Woessmann, 2009; Bai and Kung, 2011; McCleary and Pesina, 2012; Woodberry, 2012; Cantoni, 2013). These studies consider Protestant conversion as a whole. Instead of using such a binary approach, we exploit different types of missionary treatment.

Woodberry (2004, 2012) and Woodberry and Shah (2004) first document the role of Protestant missionaries on the consolidation of liberal democracy and emphasize the introduction of the printing press and newspapers as a potential mechanism. They identify the relationship between Protestantism and democracy using a cross-country identification strategy. On the contrary, we estimate the long-term effects of the printing press in sub-Saharan Africa using a variety of within-mission empirical strategies. Our econometric analysis moves beyond selection from missions's location. Our empirical work relies on the building of a new geocoded dataset and on the use of several previously unexploited historical archives.

Our results also complement a growing literature documenting the persistence of development paths (Acemoglu et al., 2001; Glaeser and Shleifer, 2002; Porta et al., 2008). Focusing on literacy and religious affiliation, Cogneau and Moradi (2014) find that colonial border effects still persist today. At a more refined level, Michalopoulos and Papaioannou (2014) shed light on the long-term growth consequences of sub-national institutions. Protestant missionaries are potential agents of change at the sub-national level. While this literature mainly compares regions with different institutional or colonial history, we highlight access to the printing press as a specific long-term determinant of civic and social capital, within regions sharing the same institutional or ethnic background.

As noticed by Feyrer and Sacerdote (2009), Huillery (2009) and Huillery (2011), historical events can explain heterogeneous development dynamics. Recent micro-oriented studies therefore isolate specific channels through which a development dynamic was durably established (Nunn, 2008; Huillery, 2009; Dell, 2010; Alesina et al., 2011; Michalopoulos and Papaioannou, 2011, 2013; Voigtländer and Voth, 2012). Of particular importance for sub-Saharan Africa are early investments. Wantchekon et al. (2012) highlight for example the durable impact of the first schools in Benin. There has been little research on the long-term consequences of the printing press. Rubin (2014) documents the link between the spread of the Protestant Reformation and the spread of the printing press in Europe at the time. Dittmar (2011) shows that European cities where printing presses were established between 1450 and 1500 grow faster between 1500 and 1600 than similar cities which were not early adopters. While Dittmar focuses on urbanization and economic change in the 16th and 17th Century Europe, we identify the effect of the early arrival of the printing press on contem-

porary local civic engagement and social capital in sub-Saharan Africa. Consistently with the hypothesis from historians of sub-Saharan Africa (Omu, 1978; Tudesq, 1995), we are the first to show empirically that proximity to the missionary press is associated with higher newspaper readership today, as well as with other proxies for social capital. Moreover, we find that contemporary education is higher in regions close to the missionary press. Finally, we are the first to investigate the extent to which different types of missionary investments affect different contemporary outcomes.

The rest of the paper is organized as follows. Section 2.2 presents a brief historical background on missionary activity in sub-Saharan Africa and the development of newspapers. Section 2.3 describes the data, in particular our new geocoded dataset of missions and discusses the determinants of missions' location and investments. In Section 1.4, we provide empirical evidence on the long-term impact of proximity to a printing press on civic and social capital nowadays. We also document the long-term relationship between mission investments in education and health and contemporary education- and health-related outcomes. We discuss extensively endogenous selection of missions into these different investments. Section 2.5 discusses the results and the channels of transmission. Section 2.6 concludes.

## 1.2 Historical background and persistence

**The introduction of the printing press** Protestant missionaries pioneered in the development of a written tradition for sub-Saharan African languages. Wherever they went, Protestants quickly formalized indigenous languages and printed Bibles and educational material in these languages. Following the principle of the *Sola Scriptura*, every Protestant should be capable of reading and interpreting the Bible (Eisenstein, 1980; Woodberry and Shah, 2004; Woodberry, 2012).<sup>2</sup> On the contrary, for the Catholic religious practice, reading the Bible is not necessary.

Protestant missionaries were thus among the first to bring the printing press to sub-Saharan Africa; on the contrary, Catholic missionaries had no role in bringing the printing

---

<sup>2</sup>“*Christian missionaries continued to set up printing presses in remote parts of the world to turn out Gospels and Psalters as had been done in Mainz four hundred years earlier*” (Eisenstein, 1980, p.158).

press. Moreover, Protestant missions facilitated the access to the printing press, acting as the intermediaries of its diffusion. For example, in South Africa, several mission societies acquired the printing press in Cape Colony between the 1820s and the 1870s. The Methodists acquired a printing press in Grahamstown in the 1830s. The Anglicans acquired presses for several stations in the eastern Cape in the 1860s and 1870s. In the later 19th and early 20th centuries, missionary societies outside the Cape Colony were also active in publishing, especially in Natal (Switzer, 1984).

Investing in printing technology was a better strategy than importing books, as transportation was long and native languages were mostly unknown in Europe. Due to technological constraints, printing presses could hardly be built in sub-Saharan Africa and had to be imported from Europe. Missionaries mainly imported hand press machines; nevertheless, importation was far from easy.<sup>3</sup> Wooden printing presses were, for instance, highly inflammable materials. In 1819, Thomas Stingfellow and Robert Godlonton, both English settlers and qualified printers, set sail for South Africa with a large crate containing a second-hand wooden press. Their “inflammable” machine was however impounded in Table Bay by the Acting Governor, calling a halt to their printing project before it even started (Gordon-Brown, 1979). Similarly, the Wesleyan missionaries in Grahamstown (South Africa) decided to import an iron printing press at the end of the 1820s. But their project was almost as complicated. When Reverend Stephen Kay arrived with the machine in 1830, the missionaries realized that certain characters required for the Xhosa language were missing from the typing range. The printing press therefore could not be used for almost three years, the time it took for the missing material to arrive to Grahamstone.

This anecdotal evidence illustrates the complexity of starting any printing activity in sub-Saharan Africa. Moran (1973) gives the rare details on the press size. The platen of one of the smallest wooden presses found in the 1820s (the *Brooke*) was 30cm×3.8cm×45cm (1ft×1½in×1ft5in) while it was 91cm×30cm (3ft×11½in) for the largest one. Iron presses were larger. For example, the size of the Columbian press introduced between 1812 and 1814 and which, according to Moran (1973), “*looks much like others of the period*”, ranges from

---

<sup>3</sup>At the time of Protestant missions’ settlement in sub-Saharan Africa, there existed three kinds of printing presses: the wooden press directly inherited from the old Gutenberg’s printing press (Clair, 1976); the iron press, the most famous one being the Stanhope press which appeared around 1800; and the more technological steam press that uses a rolling cylinder in printing to overcome excessive manual strain.

53cm×40cm (21in×16in) to 106cm×68cm (42in×27in).

Printing presses were also costly. Moran (1973) gives some information about the prices of printing activities in England at the turn of the century. At the beginning of the 19th Century, wooden presses cost around £65. When Stanhope introduced his first iron press around 1800, its price was £90. The price began to drop with competition; in 1808, however, it still ranged from £21 to £73. A few years later, at the end of the 1810s, the price of the Columbian press, another iron model, ranged from £100 to £125. In 1820 the price was still above £75. In the 1840s, the Albion press – following the Columbian Press – varied in size from Amateur (15cm×12cm or 7×5½in) to Double Royal (100cm×58cm or 40in×23in). The price of the Double Royal was £75.<sup>4</sup> In 1830, the average annual income in the United Kingdom per adult was £30, while the average annual wage was £20. The average worker (blue-collar) annual wage was £15.<sup>5</sup> According to Maddison’s historical per capita GDP series, average incomes in Africa around 1820 were about five times smaller than in the United Kingdom. In other words, a printing press cost on average twenty-five years of a worker’s wage. Using University of Illinois’ recent project “Measuring Worth”, £1 in 1800 represents between £70 and £4000 in 2010.<sup>6</sup>

Printing presses were costly and sizable and missionaries had to import them from Europe. Printers had to train apprentices and do much of the mechanical work themselves: “*the editor, printer, publisher and proprietor were all combined in one person*” (Gordon-Brown, 1979). Printing presses were often operated by settlers who had experience on a printing office in England. Specific educational investments also had to be made. Printing schools were established by Protestant missionaries for the indigenous population to acquire the specific knowledge required. In 1896, four printing schools were already active in South Africa (Cape Colony), Zanzibar and Malawi (Lake Nyasa) with close to 300 students (Church Missionary Society, 1896). As a consequence, printing presses were not available to the native population

---

<sup>4</sup>Not to refer to the prices of the cylinder machine which were used in the United States at the time. The simplest Koenig machine, the single cylinder, cost £900. A double machine cost £1,400 and the most advanced one £2,000.

<sup>5</sup>See e.g. Piketty and Zucman (2014).

<sup>6</sup><http://www.measuringworth.com> The estimation varies a lot depending on the method used. Using a simple purchasing power calculation, £1 in 1800 represents on average £62 in 2010. The difference is significantly larger using other methods: when comparing to the average worker’s wage, £1 can be compared to £1000; it can be compared to £4000 when considering the economic cost.



outside Protestant mission stations with a printing press.<sup>7</sup>

**The introduction of the first newspapers** Publishing material for Africa was central to the missionary ambitions since the commencement of their work (Fahs and Davis, 1935); mission presses not only published Christian texts but were also the first to produce all sorts of written material and newspapers in indigenous languages (Maake, 2000).

Because Protestant missionaries made printing presses available to the indigenous population, most of the first indigenous newspapers were printed and sponsored by mission centers. Woodberry (2012) qualifies conversionary Protestants as being “*a crucial catalyst initiating the development and spread (...) of newspapers*”.

The first newspaper intended for black readers, the *Umshumayeli Wendaba* (“Publishers of the News”), written in Xhosa, was published as an irregular quarterly in 1837 and printed at the Wesleyan Missionary Society in Cape Colony.<sup>8</sup> The *Iwe Irohin* (“The Newspaper”) was founded in 1859 as a publication directed by Reverend Henry Townsend from the Anglican Church missionary society in Nigeria. *Isigidimi samaXhosa* (“The Xhosa Messenger”), the first African newspaper edited by Africans, was first released in January 1876 and printed at the Lovedale Mission Press in South Africa. Eight years later, in November 1884, the English/Xhosa weekly *Imvo Zabantsundu* (“The African Opinion”) was published. It was the first black-owned newspaper in South Africa. The *Imvo Zabantsundu* was edited by John Tengo Jabavu, former editor of the *Isigidimi*, and perhaps “*the most widely known mission-educated African in Southern Africa*” at the time (Switzer and Switzer, 1979).

In regions where Protestant missions were less active, the first newspapers appeared only at the beginning of the 20th Century and no indigenous newspapers were created before World War I. Before the war, the printing presses were mostly owned by the colonial powers. The first paper in Abidjan (Ivory Coast) to be owned and edited by an African, the *Eclairneur*

---

<sup>7</sup>It is important to highlight that, while mission printing presses were not the only printing presses in the 19th century sub-Saharan Africa, there were the only presses to which the native population was given access to. Moreover, there were very few other presses. These other presses were the government presses. They were mainly used to print and copy administrative documents, and the indigenous population was not given access to these presses. As noted by Harmsen (1982), the spread of printing in sub-Saharan Africa was “*initially overwhelmingly religious*”.

<sup>8</sup>The London Missionary Society and Methodist missions also produced the earliest journals aimed at the Tswana Christian community from their stations at Kuruman and Thaba’Nchu. *Mokaeri Oa Becuana, Le Muleri Oa Mahuku* (“The Teacher of the Bechuana, the Announcer of the News”), which started in 1857, is regarded as the oldest newspaper in the Tswana language (Switzer, 1984).

*de la Cote d'Ivoire*, was first published in 1935 (Mytton, 1983).

**Persistence** The lag of more than one century in the timing of creation of the first indigenous newspapers might explain the persistent effect of the proximity to a printing press on newspaper readership today. This persistence may have worked through demand and/or supply.

On the one hand, places with printing missions – which were more likely to have newspapers – may have built up both physical (beginning with printing presses) and human capital (e.g. printing schools or journalism schools) that made it cheaper to run newspapers there nowadays. We discuss in Section 2.5 the role played by the regional development of the publishing industry. Unfortunately, we do not have information on newspaper supply at the local level nowadays in sub-Saharan Africa.<sup>9</sup> The few datasets available only provide information on the main national newspapers. In the Appendix, we provide suggestive cross-country evidence of the persistence of newspaper supply.<sup>10</sup> This evidence supports the notion of the early emergence of newspapers around mission stations and the persistence of the concentration pattern.

On the other hand, the early availability of the printing technology enabled the local development of a culture of writing and information diffusion. Newspapers take time to consolidate. In most sub-Saharan African countries, the newly-established government tried to take control of the press after independence. These nationalizations did not succeed in countries where newspapers were well established, stable and independent before colonization. In Nigeria, for instance, despite the 1966 *coup d'état*, the ensuing military regime and the development of a state-owned press, independent newspapers managed to survive. Similarly, even during the Apartheid in South Africa, the black press and anti-Apartheid white-owned presses continued to exist. *The Daily Dispatch*, the *SASO Newsletter* or *The World* regularly diffused their anti-Apartheid stances. This was not the case in former French colonies.

Reading habits exhibit strong persistence over time. A survey on newspapers made by the Lumina Foundation across Lagos, Enugu, Oyo, Edo, Kogi, Kaduna and the River States,

---

<sup>9</sup>There is no information on the number of existing newspapers at the local level, nor information on newspaper circulation at the local level.

<sup>10</sup>We also present the few data sources on sub-Saharan African newspapers that are available and that we digitize and merge together.

highlights the persistence of reading habits in southern and middle-belt Nigeria. 59% of the respondents replied that they read newspapers as a family-inherited culture (Fraser, 2008).

## 1.3 Data and missionary locations

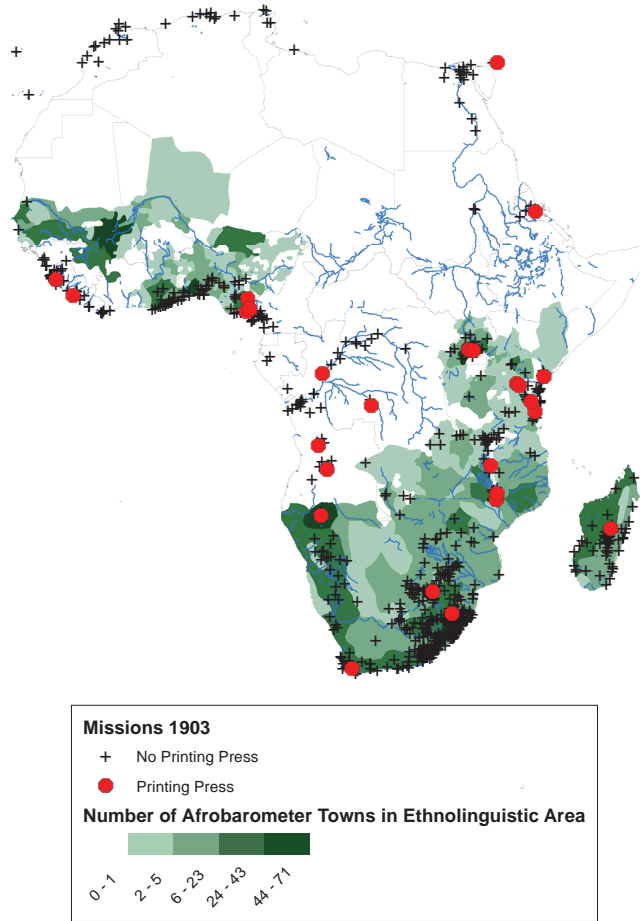
### 1.3.1 Data

#### Historical data

**Missionary activity** We construct the mission-level data from the *Geography and Atlas of Christian Missions* (Dennis et al., 1903). We geocode the maps of sub-Saharan African regions from this atlas. The maps locate all the Protestant mission stations in 1903 (an example of these maps is provided in the Appendix Figure A.2).

As opposed to other available geographic datasets of Protestant missions (Nunn, 2009), ours contains detailed information for each mission settlement. In the *Geography and Atlas of Christian Missions*, each mission station is uniquely identified in a statistical appendix providing information on the mission’s size (number of students, of missionaries, etc) and a detailed record of its activities and investments. For example, we know whether each mission had a printing press, a school, or a health facility. The exhaustive list of variables and a reproduction of one page of the statistical index are provided in the Appendix Figure A.1. We digitize and geocode this information. Our sample of sub-Saharan African missions includes a total of 723 Protestant missions out of which 27 were equipped with a printing press in 1903. Figure 1.1 shows the location of mission stations and their printing presses in 1903. This dataset is, to the extent of our knowledge, the most exhaustive record of missionary investments in the world, and it is geocoded for Africa. We hope this dataset will be of use for future research.

Since very few Catholic missions are recorded in the *Geography and Atlas of Christian Missions*, we use data from Béthune (1889) to locate them. This source does not provide information on the investments conducted. However, as we discussed in Section 2.2, Catholic missionaries had no role in bringing the printing press.



**Notes:** This map is a digitized and geocoded version of plates 14 to 18 of Dennis et al. (1903). The geocoding was conducted by the authors.

**Figure 1.1** – Mission Stations With and Without a Printing Press in 1903

**Publication records** We gather information on the number of publications printed at the mission press until 1923, twenty years after the publication of the *Geography and Atlas of Christian Missions*. The *Bibliography of Christian Literature* (Rowling and Wilson, 1923) inventories all the books and reviews, religious or not, published by the missionaries. From this source, we identify 18 presses from our original sample that had a publication record in 1923. We use this additional data in Section 2.5 (“Discussion”).

**Historical controls** The *Ethnographic Atlas* (Murdock, 1967) provides precolonial characteristics at the ethnic group-level such as initial population density. The slave trade data come from Nunn (2008). We geocode this data at the mission-level. Geocoded town-level data are from Nunn (2008).

### Contemporary data

Contemporary data on civic and social capital are from the 2005 (Round 3) and 2008 (Round 4) Afrobarometer surveys.<sup>11</sup> There are 19 sub-Saharan African countries in these surveys: 10 former English colonies (Ghana, Kenya, Lesotho, Malawi, Nigeria, South Africa, Tanzania, Uganda, Zambia and Zimbabwe), 5 French (Benin, Burkina Faso, Madagascar, Mali and Senegal), 2 German (Botswana and Namibia), 1 Portuguese (Mozambique) and Liberia.<sup>12</sup> Surveys are based on interviews conducted in the local languages from a random sample of either 1,200 or 2,400 individuals of voting age in each country. Overall, they cover approximately 47,000 individuals sampled to constitute representative groups at the national level. The data is geocoded at the district level.

The Afrobarometer provides individual-level subjective data on media consumption and civic attitudes. Our proxies for social capital are newspaper readership (whether respondents read newspapers on a monthly basis) and trust (how much to you trust other country members). As to political participation, we use three different measures: registration for elections; voting in past elections; and a proxy for political participation at the local level (“Actions

---

<sup>11</sup>Each variable is described in details in the Appendix Section A.1.2. The description provides the questions used in the Afrobarometer to construct the different variables.

<sup>12</sup>To be more specific, there are 17 countries covered in both rounds, while for Burkina Faso and Liberia we only have data for the Round 4 of the Afrobarometer. Indeed the Afrobarometer data does not have a panel structure. Different individuals – sometimes from different locations – are surveyed in the different rounds, and the questions asked in the surveys may also vary from one round to the other.

as citizen”) which captures how likely individuals are to participate in a collective action or a march.<sup>13</sup>

Newspaper readership is of importance for a number of reasons. First, it can be considered as a proxy for civic and social capital (Putnam, 2000; Guiso et al., 2004). Second, there is vast evidence that it is a strong causal determinant of political participation (see e.g. Gentzkow et al., 2011; Cagé, 2014) and voter knowledge (Snyder and Strömberg, 2010). While there is a growing questioning of the future of newspapers in the Internet era, sub-Saharan Africa is one of the few places in the world where the newspaper market is still growing. This market expands as literacy steadily increases whereas other media like television or internet require capital that most sub-Saharan Africans do not have. Moreover, as suggested by Bratton et al. (2005), it is harder for governments to control newspapers while they can control radio or television by restricting supply and imposing a government monopoly (see e.g. Cage, 2015).

We also investigate the long-term effects of early missionary investments on education- and health-related contemporary outcomes. An outcome of interest would have been literacy; unfortunately this information is not available in the Afrobarometer. We thus focus on education that we measure with a discrete variable ranging from 0 to 9, describing different levels of education from none (0) to post-graduate education (9). As to health, we use a binary variable equal to one if the respondent reports there is a clinic close to her town (“Clinic”). A better measure, capturing individual health conditions, would have been to use anthropometric indicators. However, such indicators are not available in the Afrobarometer.

Finally, the Afrobarometer provides information on a set of individual controls: age, sex and ethnicity, among others.

Table A.4 in the Appendix provides summary statistics for these variables. On average, 33% of the individuals surveyed read a newspaper at least once a month; 38% in towns located 100km away or closer to the closest mission. 42% of the individuals report to trust people in general (“Trust”), 38% in towns located close to a mission. As to political participation, the numbers reported for both registration (80%) and turnout at elections (72%) seem to be very high, but this is a well-known issue with the Afrobarometer. 24% of the individuals report that they participated in a collective action or a march.

---

<sup>13</sup>Note that this last variable may also be interpreted as a proxy for social capital, as in Acemoglu et al. (2014).

## Geographic characteristics

Finally, to control for geographic characteristics at the town and mission level, we use the *Global Agro-Ecological Zones* (GAEZ) data. The data are geocoded and provide information on annual precipitation levels, the average suitability for rainfed crops, the number of agriculture growing days per year and the malaria ecology of the land.

### 1.3.2 Determinants of mission location and investments

Before turning to the empirical analysis, we analyze the determinants of mission location. We also compare missions that invested in the printing technology and missions that did not. On average, towns in the Afrobarometer are located 140 km away from the closest mission settlement, and 439 km away from the closest mission settlement with a printing press (see Table A.3 and Figure A.5 in the Appendix for more details). As a baseline, we use the 100km threshold to define towns close to a mission, but we extensively discuss other specifications.

**Mission location** A number of factors played a role in determining the location of mission settlements. Among them are access to a clean water supply, the ability to import supplies from Europe, the abundance of a fertile soil that could be used to grow crops, and a high altitude with a temperate climate (Johnson, 1967). Moreover, mission locations exhibited a form of path-dependence. The routes of initial missionary explorers determined which parts of Africa were the best-known to Europeans, as well as the locations of the earliest mission stations from which larger networks of stations were developed. The colonial railway network is another factor that affected mission locations, as well as the slave trade (Johnson, 1967; Nunn, 2010).

These known trends are for the most part confirmed in our data. In Table 1.1 we perform a t-test on the equality of means for geographic and historical characteristics of towns located near (less than 100km) and far (more than 100km) from a historical mission settlement. As for the geographic characteristics, we find that missions locate in places with a lower prevalence of malaria and a more favorable climate. They also locate in places more suitable for agriculture (measured by the suitability for rainfed crops and the number of agricultural growing days). Moreover, they locate closer to the coast. As for the historical determinants of

mission location, slave exports are higher in places near missions. We also find that missions have a higher probability to be located near historical railway networks.

An open question in the literature is the general effect of population density. Some missionaries intentionally built missions in more remote locations where the “word of God” otherwise would not have reached; whereas other missionaries recognized the benefits associated with dense populations and targeted these groups (De Gruchy, 1999; Nunn, 2010). According to Table 1.1, regions near historical mission settlements had on average a higher population density.

In our empirical analysis we only focus on regions near historical mission settlements. Moreover, our specifications always control for all the geographic and historical characteristics described in Table 1.1.

**Location of missions with a printing press** Did missions with a printing press locate in regions with specific geographical or historical characteristics correlated with determinants of civic and social capital? Table 1.2 compares the geographic and historical characteristics of missions with and without a printing press and perform a t-test on the equality of means. Missions with a printing press were not, on average, located in more geographically favored areas. None of the geographical indicators are significantly different between the two groups, except for malaria ecology which is higher for missions that invested in the printing press.

Historical characteristics exhibit a different pattern. Missions with a printing press have more favorable historical characteristics: they are closer to historical cities and explorer routes. All our specifications control for these characteristics.

**Mission investments** Did missions with a printing press conduct different types of investments or have different characteristics? Table 1.3 presents descriptive statistics comparing investments and characteristics of missions with and without a printing press. Missions with a printing press arrived earlier and have a much higher probability of being Bible Societies.

Missions with a printing press have on average more schools, both in level and per student than missions without, as well as more teachers per student. We use various empirical strategies to check that our results are not driven by these higher investments in education. Finally, missions with a printing press have more health facilities in level. However they



**Table 1.1** – Determinants of the location of missions (Comparing towns close and towns far from Missions)

	(1)	(2)	(3)
	No Mission	Mission	Diff
<b>Geographical Characteristics</b>			
Number of Growing days (%)	317.877 (43.411)	102.281 (43.904)	215.597***
Suitability for Rainfed Crops	3.658 (0.066)	4.336 (0.066)	-0.678***
Average Precipitation per Day (over year)	4.804 (0.364)	3.337 (0.368)	1.467***
Distance 2000 city (100km)	7.166 (0.273)	8.306 (0.276)	-1.140***
Distance to the Coast (100km)	569.029 (12.820)	385.200 (12.965)	183.829***
Malaria Ecology	15.253 (0.367)	9.108 (0.371)	6.145***
Distance to mission / 100	3.048 (0.049)	0.389 (0.050)	2.659***
<b>Historical Characteristics</b>			
Slave exports (per capita)	0.685 (0.260)	2.941 (0.263)	-2.255***
Railway contact	0.995 (0.034)	0.530 (0.034)	0.465***
Explorer contact	1.525 (0.107)	2.792 (0.108)	-1.266***
Initial population density / 1000	0.232 (0.021)	0.320 (0.022)	-0.089***
Distance 1400 city (100km)	5.701 (0.168)	8.425 (0.170)	-2.724***
Distance 1800 city (100km)	8.295 (0.341)	12.095 (0.345)	-3.800***
Observations	18045	29411	

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table compares geographical and historical characteristics of town with and without missions. Towns with missions are towns located near (less than 100 km away) from a historical mission settlement. Towns without missions are towns located far (more than 100 km away) from a historical mission settlement. Column 1 presents the results for places without missions. Column 2 presents the results for places with a mission. In Column 3 we perform a t-test on the equality of means. Standard errors in parenthesis are clustered at the town-level. Variables are described in the Appendix.

**Table 1.2** – Determinants of the location of the printing press (Comparing missions with and without the printing press)

	(1) No Printing	(2) Printing	(3) Diff
<b>Geographical Characteristics</b>			
Number of Growing Days (%)	53.123	60.049	-6.926 (4.920)
Suitability for Rainfed Crops	4.992	4.750	0.242 (0.356)
Average Precipitation per Day (over year)	3.413	3.791	-0.378 (1.208)
Distance to 2000 city (100km)	2.702	2.655	0.047 (0.374)
Distance to the Coast (100km)	2.310	2.747	-0.437 (0.534)
Malaria Ecology	5.187	10.295	-5.108** (1.509)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.074	0.042	0.031 (0.107)
Railway Contact	0.237	0.107	0.129 (0.081)
Explorer Contact	0.143	0.286	-0.143** (0.069)
Initial Population Density / 100	11.803	13.760	-1.958 (8.008)
Distance to 1400 City (100km)	8.872	6.969	1.903* (0.932)
Distance to 1800 City (100km)	15.587	10.704	4.883** (1.890)
Observations	651	28	679

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Tables compares the characteristics of the places where missions with and without a printing press did locate. Column 1 presents the results for missions without a printing press. Column 2 presents the results for missions with a printing press. In Column 3 we perform a t-test on the equality of means (robust standard errors in parenthesis). Variables are described in the Appendix.

**Table 1.3** – Characteristics of and investments performed by missions with and without a printing press

	No Printing	Printing	Diff
Arrival Date	1878	1872	7* (4)
Bible Society	0.052	0.357	-0.305*** (0.051)
Number of Native Workers	3.422	1.786	1.637 (3.202)
Total Population	327.822	413.464	-85.642 (120.425)
Schools	0.261	1.643	-1.382*** (0.174)
Number of Students	321.363	397.286	-75.923 (119.386)
Schools per Student (%)	0.322	1.350	-1.028** (0.370)
Teachers per Student (%)	10.087	18.861	-8.773* (4.385)
Health Facilities	0.187	1.464	-1.277*** (0.139)
Physicians per Capita (%)	0.345	0.952	-0.607 (0.581)
Health Facilities per Capita (%)	1.521	2.302	-0.781 (1.268)
Observations	651	28	679

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The tables compares the characteristics of and the investments performed by missions with and without a printing press locate. Column 1 presents the results for missions without a printing press. Column 2 presents the results for missions with a printing press. In Column 3 we perform a t-test on the equality of means (robust standard errors in parenthesis). Variables are described in the Appendix.

do not invest more in health per capita. Our specifications always control for missions' characteristics and investments.

## 1.4 Empirical analysis

### 1.4.1 Specification and identification strategy

Let  $i$  index individuals,  $j$  index the town in which individuals live<sup>14</sup>,  $e$  index the ethnicity and  $c$  index the country.  $r$  stands for the Afrobarometer round. Standard errors are clustered at

<sup>14</sup>In the Afrobarometer, individuals are assembled by the smallest unit among villages, cities or districts. We defined this unit using the latitude and longitude provided in the Afrobarometer. We call "town" this unit in the remainder of the paper.

the town level.

Equation 1.1 describes our preferred identification equation:

$$\begin{aligned}
 y_{ijec} = & \alpha_1 \text{Distance to printing press}_j + \alpha_2 \text{Distance to school}_j + \alpha_3 \text{Distance to health}_j \\
 & + \beta_1 \text{Distance to Protestant mission}_j + \beta_2 \text{Distance to Catholic mission}_j \\
 & + \mathbf{X}'_i \boldsymbol{\beta}_2 + \mathbf{Y}'_j \boldsymbol{\beta}_3 + \mathbf{Z}'_e \boldsymbol{\beta}_4 + \delta_c + \rho_r + u_{ijecr}
 \end{aligned} \tag{1.1}$$

where  $y_{ijec}$  is our outcome of interest (described in more details below).

In our main specification, “Distance to printing press $_j$ ” is the logarithm of the distance from town  $j$  to the closest mission with a printing press. Distances are measured in kilometers; we discuss below less parametric specifications. The parameter  $\alpha_1$  is our parameter of interest. It captures the impact of the proximity to a printing press. “Distance to Protestant mission $_j$ ” is the logarithm of the distance from town  $j$  to the closest mission. The distances are computed using the geocoded information described in Figure 1.1 and Section 3.1.

We control for a large set of covariates that might determine individual behaviors today and historical mission settlement. The choice of the control set comes from Nunn (2008) and Michalopoulos and Papaioannou (2011).

The vector of individual controls  $\mathbf{X}'_i$  includes the age of the surveyed individuals, their age squared and their sex.

The vector of town-level controls  $\mathbf{Y}'_j$  includes a wide range of historical and geographical factors that may have played a role in determining both mission center locations and long-term economic development. At the town level, we control for the distance to the capital city; current and historical distance to the coast; the historical exposure to the trans-Atlantic and Indian slave trades; and the precolonial population density. We also control for the distance to the closest missionary investments that may be correlated with long-term development: schools and health facilities. Similarly, we add historical investments and geographic characteristics of the closest mission: the annual precipitation level; the suitability for rain-fed crops; the number of agricultural growing days; the number of native workers, students, teachers and physicians; the total population of the mission and the arrival date of the mission. Finally, we add the logarithm of the distance to the closest Catholic mission

1889.

Finally, the vector of ethnicity-level controls  $\mathbf{Z}'_e$  include the malaria ecology of the land; the average elevation; the share of land within 10 km of water; a binary variable equal to one if any part of the railway network was build on land historically inhabited by the ethnic group and zero otherwise; a binary variable equal to one if an European explorer traveled through land historically occupied by the ethnic group; and the distance to the closest precolonial city (cities in 1400).

**Outcomes** We begin by investigating the effects of the proximity to a printing press on different measures of civic and social capital. Our proxies for social capital are newspaper readership and trust.<sup>15</sup> We use three different measures of political participation: registration for elections; voting in past elections; and how likely individuals are to participate in a collective action or a march (“Actions as citizen”).

Finally, we also investigate the long-term effects of early missionary investments on education- and on health-related outcomes (access to a clinic).

All these variables are described in more details in above Section 2.3 as well as in the Appendix.

**Reducing the sample to individuals close to a Protestant mission** We restrict our sample to regions close to historical mission settlements (we consider towns that fall in a 100 km radius buffer around missions – we show below that our results are robust to the use of other radii). We do so because, as shown in Section 2.2, regions near historical mission settlements have on average a higher population density and better geographic conditions than regions further. Moreover, all the mission stations invested in activities, especially education, that are probably correlated with long-term development. The restrictions thus aim at correcting for possible selection in mission location.

---

<sup>15</sup>Section 4.4 also discusses the results when TV and Radio are used as outcomes

## 1.4.2 OLS Estimation

**Distance to a mission** We first estimate the long-term effects of proximity to a mission, considering both Catholic and Protestant missions, on newspaper readership, trust, education, and access to a clinic nowadays. The results are presented in Table 1.4. We find no statistically significant effect of proximity to a mission, either Protestant or Catholic. Hence, if missions have long-term effects, it seems that these effects are working through specific missionary investments.

**Distance to Protestant mission investments** In Table 1.5, we then focus on the long-term effects of a number of Protestant mission investments. As discussed above, we do not have detailed investment data for Catholic missions. Reassuringly, there is no evidence on Catholic missions importing the printing press. Table 1.5 presents the results of the OLS estimations.<sup>16</sup>

First, we find a positive and statistically significant effect of proximity to a mission with a printing press on our two proxies for social capital, newspaper readership and trust. A 1% increase in the proximity to the closest mission with a printing press is associated with a statistically significant increase in the probability of reading a newspaper of 2.5 percentage points (columns 1 and 4). Similarly, it is associated with a statistically significant increase in trust of 3.6 percentage points (columns 5 and 8). In terms of magnitude, a one-standard deviation increase in the logarithm of the proximity to the printing press increases the probability of reading newspapers by 5.3% of a standard deviation, and decreases trust by 8.1% of a standard deviation.<sup>17</sup>

A variance decomposition of the results show that the proximity to the printing press and the other covariates together explain respectively 11.8% of the total variation in newspaper readership. Of these 11.8%, 1.9 to 16.5% is explained by the distance to the printing press. The other investments performed by the missions (education- and health-related investments)

---

<sup>16</sup>All the specifications control for the distances to Catholic and Protestant missions. The coefficients are not reported in the main text for the sake of space; none of them is statistically significant. Appendix Table A.13 reports the estimated coefficients for all the covariates. All the coefficients are of the expected sign.

<sup>17</sup>Table A.14 in the online Appendix reports the standardized coefficients (beta coefficient) for all the covariates.

Table 1.4 – Distance to a mission and Contemporary outcomes, OLS estimation (100km restriction)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance to Protestant mission	News b/se	News b/se	News b/se	Trust b/se	Trust b/se	Trust b/se	Education b/se	Education b/se	Education b/se	Clinic b/se	Clinic b/se	Clinic b/se
	-0.005 (0.007)	0.001 (0.009)	-0.005 (0.007)	-0.003 (0.008)	-0.014 (0.009)	-0.004 (0.008)	0.020 (0.031)	0.062 (0.041)	0.022 (0.031)	0.015 (0.012)	-0.017 (0.017)	0.014 (0.012)
Distance to Catholic mission												
Observations	28,590	28,590	28,590	15,511	15,511	15,511	28,720	28,720	28,720	28,267	28,267	28,267
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2,213	2,213	2,213	917	917	917	2,212	2,212	2,212	2,175	2,175	2,175
R-sq	0.18	0.18	0.18	0.11	0.11	0.11	0.24	0.24	0.24	0.14	0.14	0.14

**Notes:** \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

have no long-term effects on social capital proxied by trust and newspaper readership.

When we turn to the effects of mission investments on education- and health-related contemporary outcomes, we first find that proximity to a mission with a printing press has a long-term effect on education. A 1% increase in the proximity to the closest mission with a printing press is associated with a statistically significant increase in the level of education by 10.9 to 11.5 percentage points (columns 9 and 12). This corresponds to an increase in the level of education by 6.3% of a standard deviation. Moreover, not only the printing press investment has a long-term effect on education, but also the historical missionary investments in education. We find that a 1% increase in the proximity to the closest mission with a school is associated with a statistically significant increase in the level of education by 6 to 6.7 percentage points (columns 11 and 12). Interestingly, there seems to be no effect of the mission printing press and school on contemporary health (as measured by proximity to a clinic). However, this lack of effect might be due to an imperfect measure of health (reported presence of a nearby clinic), as opposed to a measure of individual health conditions. In the remainder of the paper, we only focus on social capital and education.

**Non-parametric effect of distance** Figure 1.2 shows the locally weighted scatter-plot smoothing (LOWESS) estimations of different contemporary outcomes as a function of the distance to the printing press. These graphs show the non-parametric estimation of the expected outcomes, conditional on distance to the printing press. The figures are consistent with the hypothesis of a negative linear relationship between the outcomes and the log-distance of the printing press. The deviation from the negative linear relationships observed in Figures 1.2b is mostly located at the tail of the sample where the estimation is the least efficient.

**Democracies** Finally, we estimate the long-term effects of the different investments performed by Protestant missions on political participation. When considering measures of political participation – registration, turnout and actions – we restrict our sample of analysis to democratic countries, where political participation is easier to interpret.<sup>18</sup> Table 1.6

---

<sup>18</sup>Democracies are the countries defined as such in the Polity IV Project: [www.systemicpeace.org/polity/polity4.htm](http://www.systemicpeace.org/polity/polity4.htm).

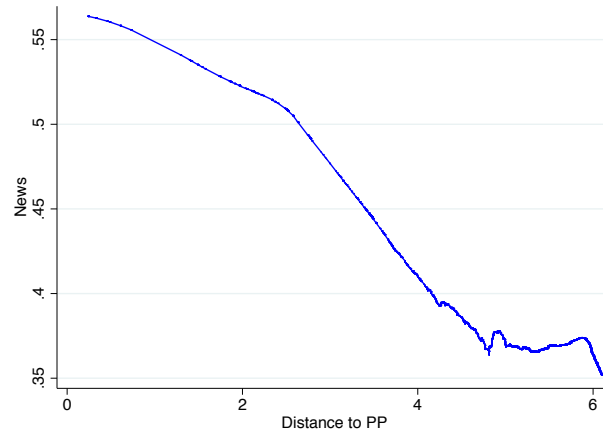


Table 1.5 – Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction)

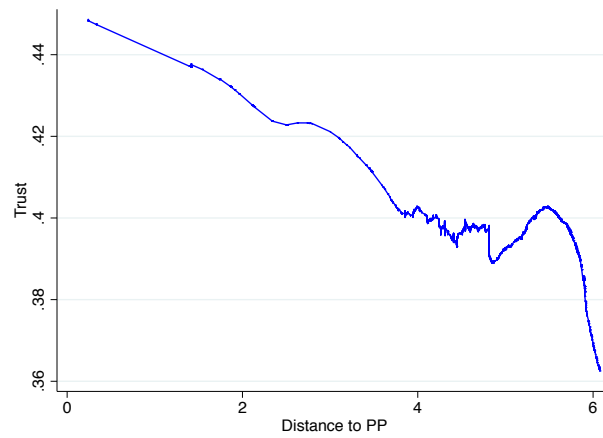
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Distance to printing press	News b/se -0.024** (0.011)	News b/se -0.002 (0.013)	News b/se 0.006 (0.009)	News b/se -0.024** (0.011)	Trust b/se -0.036*** (0.011)	Trust b/se 0.006 (0.014)	Trust b/se -0.005 (0.011)	Trust b/se -0.036*** (0.011)	Education b/se -0.105*** (0.038)	Education b/se 0.027 (0.053)	Education b/se -0.056* (0.030)	Education b/se -0.110*** (0.038)	Clinic b/se 0.005 (0.020)	Clinic b/se 0.041 (0.026)	Clinic b/se -0.021 (0.016)	Clinic b/se 0.006 (0.021)
Distance to health																
Distance to school																
Observations	28,590	28,590	28,590	28,590	15,511	15,511	15,511	15,511	28,720	28,720	28,720	28,720	28,267	28,267	28,267	28,267
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2,213	2,213	2,213	2,213	917	917	917	917	2,212	2,212	2,212	2,212	2,175	2,175	2,175	2,175
R-sq	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.24	0.24	0.24	0.24	0.14	0.14	0.14	0.14

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

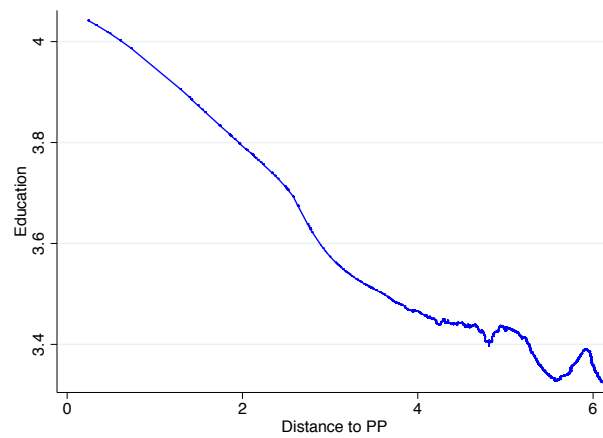
Chapter 1: The long-term consequences of the printing press



(a) Newspaper Readership



(b) Trust



(c) Education

Notes: Distance to the printing press is reported in logs.

Figure 1.2 – Effect of Distance to the Printing Press on Contemporary Outcomes. Locally Weighted Scatterplot Smoothing

presents the results.

We obtain no statistically significant effect of proximity to a printing press – nor to the other investments performed by the missions – on registration and turnout at elections nowadays. This absence of results could be explained by the well-known problem of overreporting of political participation at elections in the Afrobarometer. However, we find a positive and statistically significant effect of proximity to the printing press on the “action” variable. A 1% increase in the proximity to the closest mission with a printing press is associated with a 2.7 to 2.8 percentage points increase in the probability of taking actions as a citizen (columns 9 and 12). We find no effect on actions of the historical investment in education and health.

### **1.4.3 Endogenous selection of missions into printing and other investments**

The results suggest persistence of historical investments on contemporary outcomes, with the proximity to the printing press affecting social and civic capital as well as education, and the proximity to an historical location of a missionary school having a long-term effect on education nowadays.

Although regression (1.1) includes an extensive set of covariates, there is still the possibility of endogenous selection of missions into printing and other investments. To tackle this issue, we first develop a matching approach, and then use selection on observables to assess the bias from unobservables (Altonji et al., 2005; Oster, 2013).

#### **Matching**

As opposed to the OLS estimator that imposes linearity in the parameters, the matching estimator allows for non-linear effects of observables (Acemoglu, 2005). For the sake of simplicity, we only describe here the matching strategy with respect to selection of missions into printing, but also perform the same strategy for the health- and education- related historical investments considered in previous regressions. Our matching results, presented below, cover the three types of missionary investments.

Using a Logit model, we regress the binary variable indicating whether missions are endowed with a printing press on all the observable characteristics available at the mission

**Table 1.6** – Distance to mission investments and Political Participation, Considering only democracies, OLS estimation (100km restriction)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Turnout	Turnout	Turnout	Turnout	Registration	Registration	Registration	Registration	Actions	Actions	Actions	Actions
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Distance to printing press	0.021 (0.014)			0.022 (0.014)	-0.004 (0.016)			-0.005 (0.016)	-0.027** (0.012)			-0.026** (0.012)
Distance to health		-0.004 (0.011)		-0.004 (0.011)		-0.014 (0.011)		-0.012 (0.011)		0.008 (0.013)		0.007 (0.013)
Distance to school			-0.003 (0.008)	-0.005 (0.008)			0.012 (0.011)	0.012 (0.011)			-0.005 (0.009)	-0.003 (0.008)
Observations	12,363	12,363	12,363	12,363	5,225	5,225	5,225	5,225	11,856	11,856	11,856	11,856
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	1,188	1,188	1,188	1,188	721	721	721	721	1,188	1,188	1,188	1,188
R-sq	0.17	0.17	0.17	0.17	0.16	0.16	0.16	0.16	0.10	0.10	0.10	0.10

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary measures of political participation described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement. Only democratic countries are included in the sample.

level (these observables correspond to the variables reported in Tables 1.2 and 1.3). From this regression we compute the propensity score, which is the estimated probability of having a printing press. We then match each mission with a printing press to the mission with the closest propensity score using a one-to-one matching. The sample of missions with a printing press is called  $PP$ . Online Appendix Tables A.7 and A.8 show respectively the results of the regression from which we compute the propensity score and the balance check comparing samples  $PP$  and  $\overline{PP}$ . There are no significant differences between the two samples.

Each town in the Afrobarometer is then associated to the closest mission in the sample  $\{\overline{PP} \cup PP\}$ . We construct the treatment area as follows: we decompose the map of Africa as a Voronoi diagram using the missions in  $\{\overline{PP} \cup PP\}$  as generators (Figure ??). A Voronoi Diagram divides space into a set of regions. A set of points, the seeds, is specified beforehand and for each seed there is a corresponding region consisting of all the points closer to that seed than to any other.

We construct the following variables:

- $D\{\overline{PP} \cup PP\}_j$ : the distance from town  $j$  to the closest mission station in the sample  $\{\overline{PP} \cup PP\}$  (“Distance to printing press or likely”).
- “Treat printing press” $_j$ : a binary variable equal to one if town  $j$  is treated by a printing press (i.e. it is located in a polygon from Voronoi diagram with seed in the set  $PP$ ).

We define similarly  $D\{\overline{S} \cup S\}_j$  as the distance from town  $j$  to the closest mission station in the sample  $\{\overline{S} \cup S\}$  (“Distance to school or likely”) and  $D\{\overline{H} \cup H\}_j$  (“Distance to health or likely”). “Treat by school” $_j$  (respectively “Treat by health” $_j$ ) is a binary variable equal to one if town  $j$  is treated by an historical missionary school (similarly an historical missionary health facility).

Equation (2.2) describes the identification:

$$\begin{aligned}
 y_{ijec} = & \lambda_1 D\{\overline{PP} \cup PP\}_j + \lambda_2 \text{Treat printing press}_j \\
 & + \lambda_3 D\{\overline{PP} \cup PP\}_j \times \text{Treat printing press} \\
 & + \gamma_1 D\{\overline{S} \cup S\}_j + \gamma_2 \text{Treat school}_j + \gamma_3 D\{\overline{S} \cup S\}_j \times \text{Treat school} \quad (1.2) \\
 & + \eta_1 D\{\overline{H} \cup H\}_j + \eta_2 \text{Treat health}_j + \eta_3 D\{\overline{H} \cup H\}_j \times \text{Treat health} \\
 & + \mathbf{X}'_i \boldsymbol{\beta}_2 + \mathbf{Y}'_j \boldsymbol{\beta}_3 + \mathbf{Z}'_e \boldsymbol{\beta}_4 + \delta_c + u_{ijec}
 \end{aligned}$$

where the sets of controls  $\mathbf{X}'_i$ ,  $\mathbf{Y}'_j$  and  $\mathbf{Z}'_e$  are the same as before and  $y_{ijec}$  is our outcome of interest (contemporary newspaper readership, trust and education).

This specification aims at disentangling the effect of proximity to the printing press (respectively school and health) from the effect of proximity to a mission with similar characteristics. To ensure that “Treat by printing press” can be compared to a treatment effect of proximity to the printing press, we drop the Afrobarometer towns just at the border of each Voronoi diagram by defining a buffer around the border.

Table 1.7 presents the results of the estimation. The effect of the proximity to the closest mission in  $\{\overline{PP} \cup PP\}$  is only statistically significant if the mission did invest in the printing press. Moreover, the results are of the same order of magnitude than those we obtain with the OLS estimation. A one-standard deviation increase in the distance to the printing press decreases newspaper readership by 7.4% of a standard deviation and trust by 6.9% of a standard deviation. The effects on education are not statistically significant, but are of the expected signs.

### Using selection on observables to assess the bias from unobservables

Despite our attempts to control for observable factors, the estimates may still be biased by unobservable factors correlated with selection into printing and subsequent social capital. We follow Nunn and Wantchekon (2011) and assess the likelihood that the estimates are biased by unobservables.

A common heuristic for evaluating the robustness of a result to omitted variable bias concerns is to look at the sensitivity of the treatment effect to the inclusion of observed

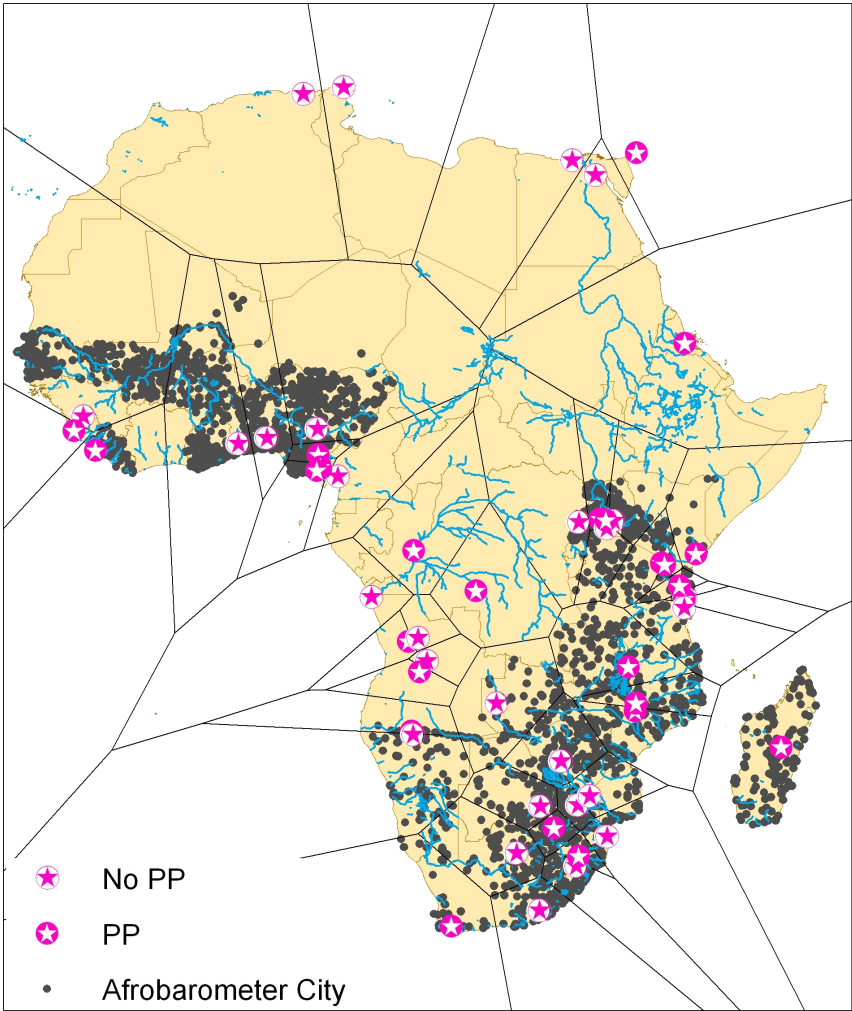


Figure 1.3 – Voronoi diagram of Africa

Notes: The seeds are the historical mission settlements with a printing press and historical mission settlements without a printing press but similar to those with a printing press

**Table 1.7** – Distance to mission investments and Contemporary outcomes, Matching estimation

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
<b>Printing press</b>			
Distance to printing press or likely	-0.022 (0.034)	0.074** (0.034)	-0.136 (0.158)
Treat printing press	0.464*** (0.162)	0.395** (0.183)	0.643 (0.848)
Distance to printing press or likely x Treat printing press	-0.067** (0.030)	-0.067** (0.033)	0.007 (0.168)
<b>Education</b>			
Distance to school or likely	-0.018 (0.021)	-0.044** (0.022)	0.007 (0.106)
Treat school	-0.058 (0.135)	-0.432*** (0.123)	0.592 (0.474)
Distance to school or likely x Treat school	0.012 (0.026)	0.072*** (0.025)	-0.130 (0.100)
<b>Health</b>			
Distance to health or likely	-0.017 (0.029)	-0.034 (0.031)	0.004 (0.139)
Treat health	-0.094 (0.114)	0.036 (0.115)	0.759 (0.577)
Distance to health or likely x Treat health	0.029 (0.022)	-0.006 (0.020)	-0.091 (0.106)
Observations	7,937	4,780	8,008
Country and Wave FE	Yes	Yes	Yes
Clusters	636	297	636
R-sq	0.21	0.10	0.30

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports matching estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects.



controls. A key issue is the need to make an assumption about the share of the outcome variance would could be explained by observed and unobserved variables together (Oster, 2013). Nunn and Wantchekon (2011) undertake direct calculations based on the theory of Altonji et al. (2005). Oster (2013) discusses their analysis and shows that this simple approximation is misleading. She makes explicit the link between coefficient movements, R-squared movements and omitted variable bias. In this paper, we follow Oster (2013)'s implementation strategy. We use the full estimator and find that it is unlikely that the entire estimated effect of the distance to the printing press is driven by unobserved variables. The results and the technical details of the estimation are presented in the Appendix Section A.3.

#### 1.4.4 Robustness checks

We perform several robustness checks. This section briefly describes them; the detailed results for these tests are available in the Appendix.

**Distance** In our main empirical specification, we reduce our sample to individuals close to a Protestant mission. Because of this specification, the effects found can be lower than the real ones if the investments performed by the missions, and in particular the printing press, have a spillover effect across regions. If the effect of the printing press vanishes in the regions outside the threshold but is present in the regions considered, then our estimates might be downward biased because of spatial externalities (Miguel and Kremer, 2004; Michalopoulos and Papaioannou, 2011). Table A.15 in the Appendix gives the results when the sample is not restricted. The effects we obtain are statistically significant and of the same order of magnitude.

We show in the Appendix that our results are robust to the use of different specifications for distance. First, Figure A.7 plots the effect of distance to the printing press from the baseline specification with different restrictions on the maximum distance to a mission. The coefficients appear to be stable across restrictions. Second, Table A.16 shows that our results are robust to an alternative specification in which the variable of interest is binary, equal to one for individuals living close to a missionary printing press (maximum 25km away from the printing press).

**Bad controls** In the main specification, we do not introduce contemporary controls that may be *bad controls* in the sense of Angrist and Pischke (2009). This issue would arise if any of the contemporary controls were also outcomes. Appendix Table A.18 presents the results of the estimation of equation (1.1) when controlling for these bad controls (contemporary religion, cash constraints, water constraints, population density in 2005). The results are not significantly different from those presented in Table 1.5.

**Mission size** Because missions that invested in the printing press might be larger, we also check that our effect is not driven by the mission size. As a falsification test, we estimate equation (1.1) with the distance to the 30 largest missions as the main explanatory variable.<sup>19</sup> The results are given in the Appendix Table A.19. Distance to the largest missions has no effect on newspaper readership nowadays.

**British colonies and legal origins** Appendix Table A.21 shows the results of the estimation of equation (1.1) for a sample restricted to former British colonies. Our results are robust to this restriction and the magnitude of the effect remains unchanged. Our results thus cannot be entirely attributed to the colonizer’s legal origins.

**Cluster** We also estimate equation (1.1) using two-way clustering at the closest mission- and town-levels. Clustering the standard errors at the mission level might be relevant should the closest mission be also considered as a treatment. Table A.20 shows that our results are robust to this two-way clustering strategy.

**Probit** Our results are also robust to using a Probit rather than an OLS specification (Appendix Table A.17).

### 1.4.5 Additional results

**Radio and television** We investigate the long-term effects of proximity to the printing press on other media outlets, and more specifically on whether individuals listen to the

---

<sup>19</sup> The size of the mission is the total number of students, missionaries and indigenous teachers and workers. Within these 30 largest missions, only two have a printing press.

newson radio or watch the news on television. The results are presented in the Appendix Table A.12. The empirical specification is the same that in Table 1.5. Whether we should expect a significant effect of proximity to a printing mission on radio and television, and the direction of the effect, is an open question. Depending on the circumstances, television and radio can indeed be either complements or substitutes to newspapers. For example, Strömberg (2004) shows that the spread of radio in 1920-1940 in the United States increased voter turnout, particularly in rural areas. This finding contrasts with Gentzkow (2006) who provides strong evidence that television caused substitution away from newspapers, leading in turn to a large drop in turnout. Both findings can be rationalized; because of its new technological feature, the radio caused less substitution away from newspapers.

We find a negative and statistically significant effect of the distance to a printing mission on listening to the news on radio and watching the news on television. The magnitude of the effect is small, however, and is only significant at the 10% level for television. A 1% increase in the distance to the closest mission with a printing press is associated with a 1.2 percentage point decrease in the probability of listening to the news on radio, and a 2.1 percentage points decrease in the probability of watching the news on television. When we turn to the standardized coefficients, we obtain that a one-standard deviation increase in the logarithm of the distance to the printing press decreases the probability of listening radio or watching television by around 4% of a standard deviation. There is no statistically significant effect of the distance to the closest mission with a school or a health facility on radio and television consumption nowadays.

**Societies** Finally, based on our historical readings of the different preferences of missionary societies for different investments, and in particular the printing press, we use missionary societies to construct an alternative proxy for mission investments.<sup>20</sup>

In the early period of missionary expansion, missionaries faced numerous difficulties despite the support of an increasing number of church ministers (Ellis, 1844). As a result of the raising popularity of the missionary work and the struggles missionaries faced, different

---

<sup>20</sup>Bai and Kung (2011), in their study of protestantism and economic prosperity in China, similarly compare the effects different Protestant denominations had on economic outcome, different denominations emphasizing Calvinist principles differently.

associations of missionaries and evangelists started to emerge in the late 18th Century. The first societies were mostly launched under the impulsion of the evangelists, for instance the Baptist Missionary Society (BMS) founded in 1792.

In the early 19th Century, numerous and diverse societies had been created. The Church Missionary Society (CMS), for instance, was founded in 1800 by the Church of England in response to the raising popularity of the missionary work (Stock, 1899). The aim of these organizations was to coordinate efforts and funds of the missionary work. Their priorities differed and depended on the preferences and means of the societies' command. The CMS, for instance, "*in the absence of missionaries, [...] fell back upon the printing press as an agent of evangelization*" from the start of its activity (Stock, 1899, p.75). On the contrary, the Christian Missionary Alliance, founded in 1887, had not invested in a single printing press ten years later, even though it had already settled 52 missions around the globe (Dennis et al., 1903).

There are 5,535 missions in the entire world outside sub-Saharan Africa (6,258 including sub-Saharan Africa) reported in the *Geography and Atlas of Christian Missions*. Only 2.3% of these missions had a printing press.<sup>21</sup> These missions were affiliated with 262 different societies among which 69 are present in both sub-Saharan Africa and the rest of the world.

Let "Society  $PP_m$ " be the share of missions from mission  $m$ 's society equipped with (at least) a printing press in all the regions of the world *outside* sub-Saharan Africa. In sub-Saharan Africa, we know the originating society for 679 out of our 723 missions (44 missions were either affiliated to a society existing only in sub-Saharan Africa or not affiliated to a society). Among the 69 originating societies recorded, 16.8% had a printing press in at least one of their missions. Assuming that the larger this share, the more likely it was for a mission associated with this society to invest in a printing press in sub-Saharan Africa, we use the society as an alternative proxy for missions' investments.

Appendix Table A.11 presents the results. The empirical specification is the same that in Table 1.5 but the explanatory variable is now the distance to the closest mission from a society with at least a printing press outside SSA. We find that this distance has a negative and statistically significant effect on newspaper readership nowadays. The effect is only

---

<sup>21</sup>This is slightly lower than in sub-Saharan Africa where the percentage of the missions with a printing press was 3.7%.

statistically significant at the 10% level.

We have shown that areas close to missions with printing presses in 1903 have higher civic and social capital today. These effects seem specific to missions with printing presses; they are not related to other observed mission characteristics. One can interpret these results as a suggestive evidence of a causal effect of printing presses on newspaper readership and social capital nowadays. In the next section, we discuss the possible channels that may explain these persistent effects.

## 1.5 Discussion

The long-term effects of printing missions on civic and social capital nowadays seem specific to missions with presses and is not related to other observed mission characteristics. However, a caveat of our analysis is that we cannot separate the effect of printing per se from other mission characteristics that might be associated with it. In particular, determining whether the persistence of the effect works through the supply side or the demand side is particularly challenging empirically.

On the one hand, it may be that places with printing missions were more likely to have newspapers, and that these places also built up physical and human capital that made it cheaper to run newspapers there today. A limitation of our dataset is that we do not know the location of current and historical newspapers nor the one of current printing presses. We circumvent this limitation by using alternative historical sources. Using Rowling and Wilson (1923)'s inventory of all the books and reviews published by the missionaries, we gather information on the number of publications printed at each mission press until 1923. In the publication record data, we find that 18 missions with a printing press (out of the 27 missions equipped with a printing press in 1903) had a publication record in 1923. These printing presses were probably the most productive or the ones with the most stable production across time.

The missions where these printing presses were located did not exhibit any particular type of geographical or historical characteristics, nor did they specialise in any kind of specific investments (Appendix Tables A.5 and A.6). We therefore interpret the presence of such

records as a proxy for the activity of the printing presses: printing presses with a publication record in 1923 were probably the most active. Table 1.8 shows the result of regression (1.1) when we add distance to the closest mission with a publication record in 1923 as a control. The outcomes of interest are as before newspaper readership, trust and education. Columns (1), (5) and (9) report the corresponding OLS estimates from Table 1.5 for convenience.

We obtain a negative and statistically significant effect of the distance to a printing press with a publication record in 1923 on newspaper readership nowadays. A 1% increase in the proximity to the closest mission with a printing press and a publication record increases the probability of reading a newspaper by 5.9 to 7.8 percentage points (columns 2 to 4). Moreover, once we control for the distance to a mission with a printing press and a publication record, the coefficient associated with the distance to a printing press is no longer statistically significant, whether or not we control for other missionary investments (columns 3 and 4). This suggests that the impact of the distance to a printing press on newspaper readership may be explained by the development of publishing activities. However, the proximity to a printing press and a publication record does not entirely capture the effect of the printing press on contemporary trust (columns 5 to 8).<sup>22</sup>

The results of Table 1.8 have to be taken carefully since they rely heavily on the only archival source we found to track the middle-run evolution of the publishing activities of missions with a printing press. They are in line with our assumption that the long-term impact of the printing press on newspaper readership may be due to the early development of the newspaper industry around missions with a printing press. But we cannot ultimately separate the effect of printing per se from other mission characteristics that might be associated with it, nor net out the supply and the demand effects that may explain persistence.

On the other hand, the early availability of the printing technology may have enabled the local development of a culture of writing and information diffusion. This would be consistent with Dittmar (2011) who, building on Habermas (1989), argues that proximity to the printing press has enabled the development of a culture of information exchange and of an urban, bourgeois public sphere. Moreover, as we highlighted above, reading habits exhibit

---

<sup>22</sup>In the Appendix Tables A.22 and A.23, we show respectively that these results are robust to clustering the standard errors at the closest mission with a publication record level and to using two-way clustering at the closest mission with a publication record and the village levels.

strong persistence over time. In other words, the historical printing press may have shifted preferences for reading.

To summarize, our analysis based on publication records suggests that the printing press may have changed the supply side of the printing industry. Similarly, the literature has emphasized the demand-changing effects of the printing press. Finally, a dynamic affecting both the supply and demand channels appears to be the most probable scenario to explain the persistence of the effects.

## 1.6 Conclusion

In this paper, we study the impact of the early introduction of the printing press by Protestant missionaries in the 19th Century. The evidence we obtain from a variety of identification strategies is consistent with our hypothesis that the early introduction of the printing press has long-term effects on contemporary civic and social capital. A wide literature points to higher social capital leading to higher economic activity and higher-well-being. Our results suggest that a better understanding of African media development and of the determinants of social capital at the sub-national level will be key for the future of African democracy and economic change.

**Table 1.8** – Proximity to a printing Press and Contemporary outcomes: Effect of the regional development of the publishing industry (1923), OLS estimation (100km restriction)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	News	News	News	News	Trust	Trust	Trust	Trust	Education	Education	Education	Education
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Distance to printing press	-0.024** (0.011)		0.021 (0.020)	0.022 (0.020)	-0.036*** (0.011)		-0.027* (0.016)	-0.027* (0.016)	-0.105*** (0.038)		-0.059 (0.063)	-0.064 (0.063)
Distance publication record		-0.059*** (0.013)	-0.078*** (0.024)	-0.078*** (0.024)		-0.038*** (0.014)	-0.015 (0.021)	-0.015 (0.021)		-0.130*** (0.049)	-0.079 (0.080)	-0.080 (0.079)
Distance to health				-0.004 (0.012)				0.002 (0.013)				0.011 (0.052)
Distance to school				0.003 (0.009)				-0.007 (0.010)				-0.062** (0.030)
Observations	28,590	28,590	28,590	28,590	15,511	15,511	15,511	15,511	28,720	28,720	28,720	28,720
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2,213	2,213	2,213	2,213	917	917	917	917	2,212	2,212	2,212	2,212
R-sq	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.24	0.24	0.24	0.24

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.





## Chapter 2

### Sex and the Mission

The conflicting effects of early

Christian investments on sub-Saharan

Africa's HIV epidemic

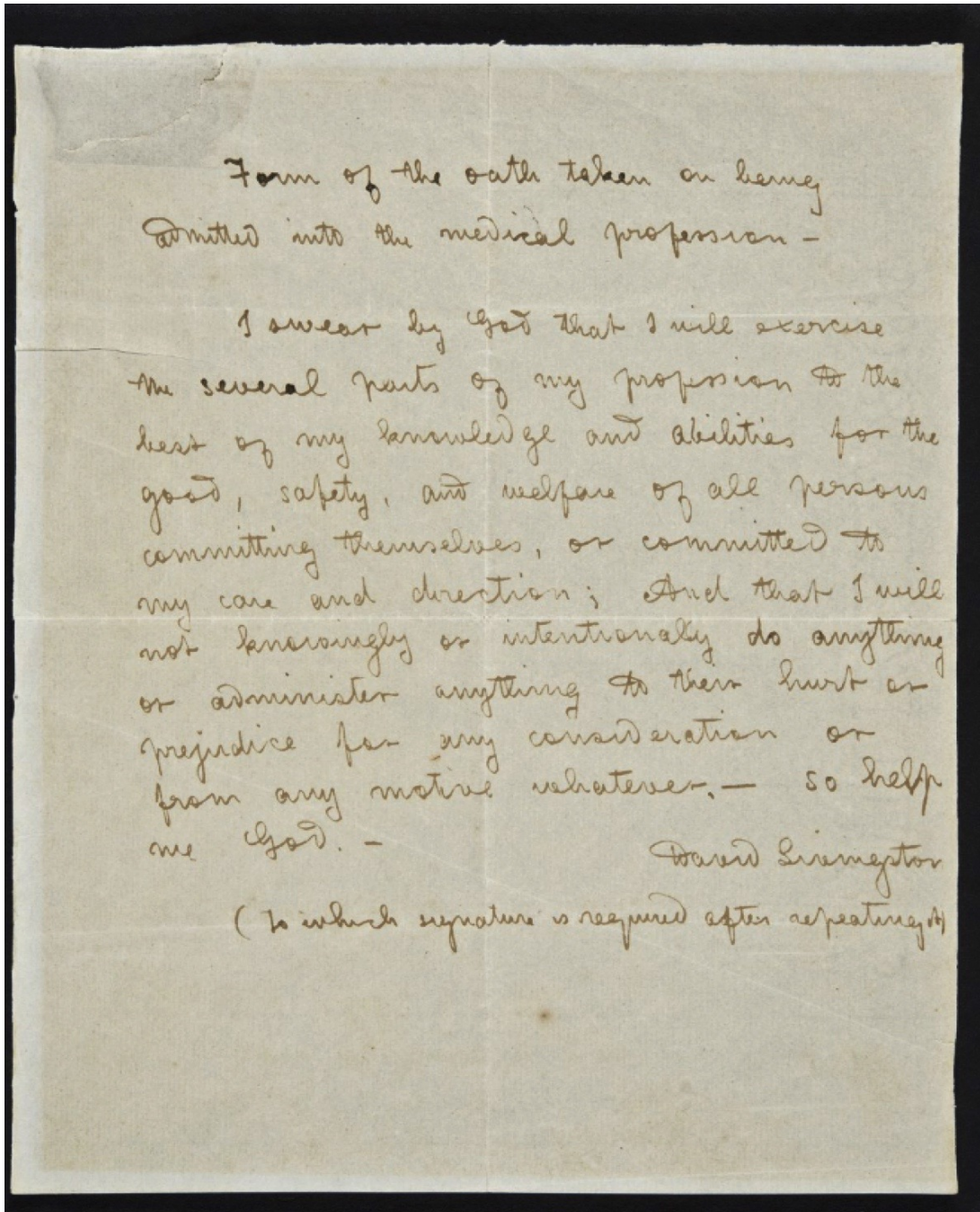


Figure 2.1 – David Livingstone’s Medical Oath

**Note:** This oath was written in 1840, the time at which Livingstone was pursuing his medical studies in London and when he was forming the idea of leaving for South Africa. He would arrive to Cape Town around one year later in 1841.

**Transcription:** “Form of the oath taken on being admitted into the medical profession - // I swear by God that I will exercise // the several parts of my profession to the // best of my knowledge and abilities for the // good, safety, and welfare of all persons // committing themselves, or committed to // my care and direction; And that I will // not knowingly or intentionally do anything // or administer anything to their hurt or // prejudice for any consideration or // from any motive whatever. - so help // me God. - // David Livingstone // (To which signature is required after repeating it” The document comes from the Digital Catalogue of “Livingstone Online”. Creative Commons Attribution-Non Commercial 3.0 Unported

## 2.1 Introduction

This paper investigates the role played by long-term historical determinants to explain HIV/AIDS prevalence in sub-Saharan Africa. More precisely, we focus on the role played by early missionary investments.

The prevalence of HIV/AIDS is very high in sub-Saharan Africa. It also varies strongly from one country to the other (Figure 2.2). In 2012, countries like Mali and Senegal had an estimated prevalence rate of around 1%; at the other extreme, this rate was above 16% in Namibia and South Africa. We show that historical investments can be an explanatory factor for this heterogeneity.

The history of modern medicine in sub-Saharan Africa is intrinsically linked to the development of missionary activity. Missionaries were the first and only ones to invest in modern medicine there until the middle of the twentieth century (Vaughan, 1991, Ch.3). David Livingstone, the nineteenth century pioneer missionary, explorer, and slave abolitionist, was a doctor himself. His medical oath (Figure 2.1), taken just months before he first set sail for Cape Town, reads almost like a prayer. This rhetoric is illustrative of a time when the medical professions, the healing of the body, could not be thought as separate from the mission of evangelization, the healing of the soul. This excerpt from an encounter between a mission doctor and a traditional Africa healer, take from Vaughan (1991, p.72) is also illustrative of the proximity between missionaries' medical provision and their evangelization work:

“I asked him [the traditional healer] if he had any medicine that would make a man live forever. He fumbled in his basket, drew out one of his bundles and undid it with great deliberation. Inside he showed me some earth, scraped from the surface of a cliff on a river bank, and told me that it should be mixed with castor oil and used for anointing the body, so that the body may prosper. ‘But even so’, he added pathetically, ‘we still die’. I told him that WE HAVE the Medicine for immortality.(...) it was to be had in Church.

According to the World Missionary Atlas (Beach and Fahs, 1925), there were 150 missionary physicians in Africa in 1925 and more than 235 nurses working with nearly 500 trained native nurses in 116 hospitals and 366 dispensaries. Historical investments are likely to be

associated with better development outcomes, and the development of a health infrastructure is likely to improve health provision in the long run (Huillery, 2011; Cage and Rueda, 2016). However, what happens when the provision of health is not impartial to the diseases it should aim at curing? Catholic and Protestant churches have a long history of condemning STIs as “diseases of immorality” (Vaughan, 1991). This paper investigates whether the Christian early health investments have long-term effects on the propagation of the HIV/AIDS epidemic, and how this effect varies with conversion to Christianity.

Health investments are not the only way through which missionary activity may affect the propagation of the HIV/AIDS epidemic nowadays. In particular, the Christian influence still affects the contemporary curricula of HIV/AIDS prevention taught in sub-Saharan African schools. This curriculum focuses almost exclusively on abstinence and monogamy to the detriment of the mention of the importance of using condoms (Duflo et al., 2015). These restrictive messages may explain why education is not systematically associated with lower prevalence of sexually transmitted infections. According to Duflo et al. (2015), “*just over one third of sexually active unmarried 15-19 year old girls in Kenya had ever used a male condom as of 2008-2009*”. We thus disentangle this Christian influence effect from the effect of health investments on the prevalence of HIV/AIDS.

Using contemporary individual-level data from the Demographic and Health Survey (DHS), we show that regions close to historical mission settlements are *more likely* to have a high HIV prevalence today. However, among regions historically close to missionary settlements proximity to a health investment (a hospital or a dispensary), is associated with *lower* prevalence rates.

Our econometric analysis attempts at moving beyond two forms of selection. First, historical and geographical characteristics might have determined mission station location, preventing the comparison between regions close and far from these settlements. Protestant missionaries did choose to locate in geographically favored areas (Johnson, 1967; Nunn, 2010). Second, among settlements, there is heterogeneity in the activities undertaken by missionaries such as printing, health and education. There may be endogenous selection of missions into these activities.

To address the first form of selection, the one from missions’ location, we follow Cage

and Rueda (2016) and restrict our sample to regions near historical mission settlements. Because regions near Protestant missions shared similar geographic, institutional and cultural environments, this restriction isolates the specific effect of the investments from other possible long-term determinants of civic and social capital embedded in specific mission locations.

To address selection of missions into different investments, we first control for observable covariates. The set of observable covariates include geographic and historical characteristics as well as distance to historical mission stations that invested in health, educational facilities and printing presses. We then develop a matching strategy that aims at isolating the effect of proximity to a mission with a specific investment (printing press, school or hospital) from the effect of proximity to a mission with similar characteristics, but without this specific investment. The results are robust to this approach and of similar order of magnitude. Finally, we assess the bias due to unobservables using the sensitivity of the treatment to added controls.

Finally, beyond these selections issues, the main empirical challenge is to disentangle the effect of the health investments on the one hand from the one of the focus on abstinence on the other hand.

To disentangle these two effects, we investigate whether other missionary investments that missionaries used as vectors of conversion, like schools or the printing press, had any effect on the prevalence of the infection. Second, we use malaria as a counterfactual. Contrary to HIV/AIDS, malaria is *not* a sexually transmitted infection and its prevalence should not be affected by the abstinence-focused curriculum imposed by the Christian influence. Finally, we focus on individual's sexual behavior and investigate the extent to which it is affected by the Christian influence.

**Related literature** There is a growing Economic literature on the HIV/AIDS epidemic in sub-Saharan Africa. A number of papers aim at understanding differences in prevalence. Oster (2005) shows theoretically that differences within Africa can be attributed to differences in sexual behavior (e.g. the number of sexual partners), male circumcision, and epidemic timing. Oster (2012) also provides empirical evidence of a positive relationship between exports and new HIV infections. In this paper, we control extensively for those determinants

that have been highlighted in the existing literature. In particular, we combine our data on HIV/AIDS prevalence with data on African roads and investigate at the micro-level the effect of road density of HIV/AIDS prevalence.

We contribute to this literature by investigating the role played by historical determinants. Bertocchi and Dimico (2015) also study the long-term determinants of the high rates of HIV infection in sub-Saharan Africa. Their focus is on family structure and sexual behavior as shaped by the demographic shock following the transatlantic slave trade. On the contrary, we focus on the long-term impact of missionary investments. To the extent of our knowledge, Mantovanelli (2015) is the only paper also considering the impact of missionary activities on HIV/AIDS and to highlight the importance of their impact on moral values and culture. However his focus is on missions' location and he does not take into account the different investments performed by the missions. In particular, he does not capture the positive long-term impact of missions through the early health investments.

Our paper is also related to the randomized experiments in the literature that highlight the limits of HIV prevention education focused on abstinence before marriage to reduce the HIV/AIDS prevalence (Dupas, 2011; Duflo et al., 2015).<sup>1</sup> Consistently with the findings of this literature, we show that the Christian influence – independently from the health investments – has a long-term positive impact on prevalence. Moreover, we document the role played by the use of condoms.<sup>2</sup>

More generally, there is a growing literature on HIV prevention. Della Vigna and La Ferrara (2015) review the impact of media exposure. Banerjee, La Ferrara and Orozco (2015) implemented a randomized control trial to estimate the impact of a TV show on risky sexual behavior, gender norms and domestic violence in Nigeria.

Finally, our article is more broadly related to the literature on the long-term determinants of development of missionary activity. The early work of Protestant missionaries has been largely associated with improving development in the long-term, through an increase in social capital (Woodberry, 2012; Cage and Rueda, 2016), human capital (Gallego and Woodberry, 2010; Valencia-Caicedo, 2014), or gender equality (Nunn, 2014). Here, we consider both

---

<sup>1</sup>See also Baird et al. (2012) who find that in Malawi, monthly cash transfers to the families of out-of-school girls significantly reduced HIV infection rates after 18 months.

<sup>2</sup>Thornton (2008) finds that when people learn they are HIV positive they increase their purchase of condoms, by only by about one condom.

Protestant and Catholic missionaries and show that missionaries have both a negative and positive impact on long-term health.

The rest of the paper is organized as follows. Section 2.2 presents a brief historical background on missionary activity in sub-Saharan Africa, the early investments in modern medicine and the focus of a number of religious groups on abstinence before marriage. Section 2.3 describes the data and presents summary statistics. In Section 2.4, we provide empirical evidence on the long-term impact of missionary activity on HIV/AIDS prevalence nowadays, and disentangle between the positive effect of health investments and the negative effect of the focus on abstinence. Section 2.5 discusses our results and presents a number of robustness checks. Section 2.6 concludes.

## 2.2 Historical background

### 2.2.1 Missions and early health investments

The history of modern medicine in sub-Saharan Africa is closely linked to the development of missionary activity. Missionaries were the first and only ones to provide this form of healthcare until the middle of the twentieth century (Vaughan, 1991, Ch.3). Missionary work was often conducted by doctors or nurses; David Livingstone, who is probably the best known figure in the history of African missionaries, was a doctor himself. In his *Missionary Travels and Researches in South Africa*, the accounts of his exploratory travels in South Africa, he describes his view of missionary activity as “much more than it is implied in the usual picture of a missionary, namely, a man going about with a Bible under his arm.” (Livingstone, 1858, p.857). Although Livingstone’s ultimate goal is the promotion of trade with the local tribes as means to improve development and end the slave trade, he emphasizes the crucial role of health and education because “neither civilization nor Christianity can be promoted alone. In fact, they are inseparable.”

From the end of the nineteenth century to the early twentieth century, there was a rapid increase in the activity of medical work among missionaries. The Edinburgh Medical Missionary Association was founded in 1841, and in 1891 the Church Mission Society established



a new “Medical Committee” that was specifically in charge of managing the medical work of the society.

The early Christian provision of health care persisted after the colonization and is particularly influential in the design of health care in poor countries (Idler, 2014). In 1968 the Christian Medical Commission (CMC) was established, it institutionalized the long-lasting tradition of Christian medical and healing work around the world. The importance of such institution is recognized by the World Health Organization that has designed the agenda on health care provision in poor countries in conjunction with the CMC and the organizations that replaced it after its dissolution in 1998 (Idler, 2014). According to Pr. Dan Kaseje’s address for the WHO on the contribution of the CMC to health care in Africa, policies such as the essential drugs initiative in 1987 (or Bamako initiative), that aimed at “solving the problems in the financing of primary health care” and lead to the commitment of African health ministers to ensure a regular supply of drugs were put together by the WHO and the UNICEF following the guidelines of the CMC.<sup>3</sup>

### **2.2.2 The Christian influence and the focus on abstinence**

Missions have invested in different activities in Africa, health, but also education or the printing press. The main role of the missions was to convert the local population to the Christian faith. Medical activity was just one way to achieve this goal, because the notion of healing the body and soul is so deeply rooted in the Christian dogma. The Gospel of Matthew 10:1 reads: *“Jesus called his twelve disciples to him and gave them authority to drive out impure spirits and to heal every disease and sickness”*. Medical work, as this form of *“healing authority”*, was *“part of a programme of social and moral engineering through which Africa would be saved”* (Hardiman, 2006).

Can a medicine embedded in Christian beliefs efficiently treat diseases to which the Christian dogma is not impartial? This question emerged since the colonial times. In Uganda, there was a large increase in the prevalence of syphilis from the mid nineteenth century, which led to one of the first public health campaigns launched in sub-Saharan Africa by the

---

<sup>3</sup>The notes of the speech can be found online. They provide many more examples of the projects that have been put together by the CMC and the WHO

colonial powers (Setel et al., 1999, p.99). The priority of the campaign, that was agreed between European missionaries, colonial administrators, and the African chiefs, was not the treatment of patients or the diffusion of scientifically accurate information, but to fight the innate sinfulness of the African society Vaughan (1991). This moral combat was in turn supposed to eradicate the *endwadde ez'obukaba*, the “disease of immorality”.

The public campaign against syphilis in Uganda is illustrative of the tension between medical work and the Christian morality in the early twentieth century. Today, in the campaigns against HIV/AIDS, religious groups favor prevention messages focusing on abstinence before marriage, rather than condom use. These groups are also heavily involved in the design of sex-education curricula delivered in many public schools. According to Duflo et al. (2015), in Kenya as well as in many other sub-Saharan African countries, the resulting curriculum teaches the biology of AIDS and HIV transmission, how to care for people living with AIDS, and prevention. The prevention component stresses abstinence before marriage, followed by faithfulness in marriage as the most effective way to prevent sexually transmitted infections. It also teaches skills such as saying no and resisting peer-pressure. The official textbook does not mention condoms nor contraception.<sup>4</sup>

Yet, it has been shown in the literature that HIV curriculum stressing abstinence before marriage does not reduce pregnancy or STIs. On the contrary, a sex-education programs that do explain the use of condoms are effective in the reduction of STIs (Dupas, 2011; Duflo et al., 2015). This article shows that the historical provision of public goods by Christian mission has conflicting effects on HIV/AIDS prevalence that reflect the problematic approach of Christianity to STIs. On the one hand, historical Christian health investments are associated with lower prevalence of HIV/AIDS. On the other hand, Christian attitudes towards contraception and condom use are associated with higher prevalence rates. These values also tend to be stronger around certain historical missions.

---

<sup>4</sup>“*The curriculum does not mention condoms and provides only limited scope for teachers to discuss protected sex in response to students’ questions. It does not cover partner selection, and although they cover love relationships between same)age boys and girls, the official textbooks do not mention cross-generational relationships (and their associated risk)*” (Dupas, 2011).

## 2.3 Data Sources and Description

### 2.3.1 Contemporary Data

#### Demographic and Health Survey

We use the Demographic and Health Survey (DHS) to estimate the HIV prevalence at the subnational level, as well as the contemporary characteristics of the regions of interest.

The DHS are standardized nationally representative surveys in developing countries. Women and men answer questions related to their anthropometrics, living conditions, health, sexual reproduction behaviors, and attitudes and beliefs towards different health outcomes, in particular towards the HIV epidemic.

Since the 2000s, the DHS also collect biomarker data on STIs. Biomarkers are objective physical or biologic measures of health conditions. Using field-friendly technologies, the DHS Program can now to collect biomarker data relating to a wide range of health conditions, including infectious and STIs, chronic illnesses (such as diabetes, micronutrient deficiencies), and exposure to environmental toxins.<sup>5</sup>

We use the DHS data for sub-Saharan African countries where both GPS information and biomarker data on HIV are provided. Figure 2.4 shows the locations of these clusters. There are a total of 13307 clusters distributed among 20 countries: Burkina Faso, Burundi, the Democratic Republic of Congo, Côte d’Ivoire, Cameroon, Ethiopia, Ghana, Guinea, Kenya, Liberia, Lesotho, Mali, Malawi, Mozambique, Rwanda, Senegal, Swaziland, Uganda, Zambia, and Zimbabwe. The information is collected in several rounds, from 2003 to 2014. Unfortunately, not all the countries are systematically surveyed so our resulting pseudo-panel is extremely unbalanced. As a result, our variable of interest is HIV prevalence (a stock), rather than incidence (the flow, number of new cases). As an illustration, consider the HIV prevalence at  $t$ ,  $h_t$ , as the number of people aged 15 to 49 years old living with HIV or AIDS in a given region. Denote  $d_t$  the number of deaths from HIV/AIDS from  $t - 1$  to  $t$ ; then the

---

<sup>5</sup>Dried blood spots (DBS) on filter paper are increasingly used for HIV testing. This method of specimen collection has eliminated the need for cold chain and/or refrigeration of specimens, reducing considerably the complexity of storage in the field and transport to the laboratory. For more information, visit the DHS website.

incidence  $in_t$  is

$$in_t = h_t - h_{t-1} + d_t$$

Normalizing by the population size to obtain rates, the incidence rates tells how likely it is to get the disease in a given region (or how fast the epidemic is moving), whereas prevalence rates indicate how widespread the disease is. Incidence rate requires yearly data on prevalence and mortality form HIV/AIDS. Previous research that focuses on incidence rates typically uses country-level estimates (Oster, 2010, 2012).

Prevalence rates are an imperfect measure of the risk of contracting the disease, but they capture the how the disease has accumulated. Because we are interested in historical determinants of the infection’s spread, as opposed to public policies in reaction to it, measuring the stock of the disease is very informative as well.

### Geographical data

Following Michalopoulos and Papaioannou (2013, 2014), we use light density data as a proxy for urbanization, population density, and regional GDP. To control for geographic characteristics at the town and mission-level, we use the *Global Agro-Ecological Zones* (GAEZ) data compiled by the Food and Agriculture Organization (FAO). From this source, we extract precipitation levels, number of growing days per year, malaria ecology, and terrain ruggedness.

Finally, we also use data on malaria incidence over time to compare to our results for HIV prevalence. The raster maps of malaria incidence in Africa from 2000 to 2011 come from the Malaria Atlas Project (Bhatt et al., 2015, MAP). The MAP provides geographical information on malaria incidence, use of Insecticide Treated Nets and other preventive treatment. It is compiled using multiple data sources (among them, surveys conducted across millions of households) and it is extrapolated geographically using a spatio-temporal Bayesian geostatistical model across 27573 georeferenced population clusters.

The data of African roads comes from the NASA Socio Economic Data and Application Center. The data set combines the best available roads data by country into a global roads coverage, using the UN Spatial Data Infrastructure Transport (UNSDI-T) version 2 as a common data model. The date range for road network representations ranges from the 1980s

to 2010 depending on the country. Updated data for 27 countries and 6 smaller geographic entities were assembled by Columbia University's Center for International Earth Science Information Network (CIESIN), with a focus largely on developing countries with the poorest data coverage (Socio-Economic Data and Application Center (SEDAC), 2016)

## 2.3.2 Historical Data

### Missionary Investments

We originally compiled the data of missionary investments and their locations in a previous article (Cage and Rueda, 2016).

We construct the mission-level data from the *Geography and Atlas of Christian Missions* (Beach, 1903). We geocode the maps of sub-Saharan African regions from this atlas. The maps locate all the Protestant mission stations in 1903 (an example of these maps is provided in the online Appendix of Cage and Rueda (2016)).

As opposed to other available geographic datasets of Protestant missions (Nunn, 2014), ours contains detailed information for each mission settlement. In the *Geography and Atlas of Christian Missions*, each mission station is uniquely identified in a statistical appendix providing information on the mission's size (number of students, of missionaries, etc) and a detailed record of its activities and investments. For example, we know whether each mission had a printing press, a school, or a health facility. The exhaustive list of variables and a reproduction of one page of the statistical index are provided in the online Appendix to Cage and Rueda (2016). Our sample of sub-Saharan African missions includes a total of 723 Protestant missions out of which 27 were equipped with a printing press, 99 had a dispensary, 38 had a hospital, and 86 had a school (a high school or a boarding school, by default, most missionary stations conducted literacy work). Figure 2.5 shows the locations of the missions and their investments.

### Additional Historical Controls

The *Ethnographic Atlas* (Murdock, 1967) provides precolonial characteristics at the ethnic group-level, such as pre-colonial population density. Data on ethnic-level slave trade come

from Nunn (2008). Georeferenced Data on historical population density is from the History Database of the Global Environment (HYDE), computed by the Dutch Environmental Assessment Agency. We geocode these data at the mission and DHS cluster levels.

### 2.3.3 Determinants of mission location and investments

Johnson (1967), Nunn (2014), and Cage and Rueda (2016) discuss extensively the determinants of missionary location. Because of potential selection in missionary location, our empirical analysis restricts the focus to regions near historical mission settlements. Moreover, our specifications always control for all the geographic and historical characteristics that may have influenced location choice, such as the malaria ecology, proximity to the coast, suitability for rainfed crops, and proximity to historical rail lines and explorer routes.

Before turning to the empirical analysis, we analyze the determinants of missionary investments. We compare missions depending on whether they invested in health facilities (Table 2.1) and whether they had a printing press (Table 2.2). We choose these two investments because they are those that we will discuss the most in the empirical analysis.

Table 2.1 compares the geographic and historical characteristics of missions that did invest in health with those that did not. Missions that invested in health facilities were not, on average, located in more geographically favored areas. None of the geographical indicators are significantly different between the two groups, except for malaria ecology which is higher for missions that worked in health. Table 2.2 shows that the same results hold for missions with a printing press.

Historical characteristics exhibit a different pattern. Missions working in health and those with a printing press have more favorable historical characteristics: they are closer to historical cities and explorer routes. All our specifications control for these characteristics.

**Table 2.1** – Summary Statistics: Characteristics of Missions with Health Investments

	(1)	(2)	(3)
	No Investment	Investment	Diff
<b>Geographical Characteristics</b>			
Number of Growing Days (%)	53.674	51.990	1.684 (2.711)
Suitability for Rainfed Crops	4.927	5.286	-0.359* (0.196)
Average Precipitation per Day (over year)	3.532	2.861	0.671 (0.664)
Distance to 2000 city (100km)	2.735	2.511	0.224 (0.205)
Distance to the Coast (100km)	2.335	2.288	0.047 (0.293)
Malaria Ecology	4.910	8.032	-3.123*** (0.825)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.067	0.100	-0.033 (0.058)
Railway Contact	0.243	0.170	0.073 (0.045)
Explorer Contact	0.138	0.208	-0.070* (0.038)
Initial Population Density / 100	10.442	19.676	-9.233** (4.373)
Distance to 1400 City (100km)	9.235	6.409	2.825*** (0.500)
Distance to 1800 City (100km)	16.386	9.977	6.409*** (1.011)
Observations	573	106	679

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  The table reports summary statistics. Column (1) reports average values for missions without a health investment. Column (2) reports average values for missions with a health investment. Column (3) reports the difference between columns (1) and (2) and the significance of a standard t-test on mean differences. Standard errors are in parenthesis.

**Table 2.2** – Summary Statistics: Characteristics of Missions with the Printing Press Investments

	(1)	(2)	(3)
	No Investment	Investment	Diff
<b>Geographical Characteristics</b>			
Number of Growing Days (%)	53.123	60.049	-6.926 (4.920)
Suitability for Rainfed Crops	4.992	4.750	0.242 (0.356)
Average Precipitation per Day (over year)	3.413	3.791	-0.378 (1.208)
Distance to 2000 city (100km)	2.702	2.655	0.047 (0.374)
Distance to the Coast (100km)	2.310	2.747	-0.437 (0.534)
Malaria Ecology	5.187	10.295	-5.108** (1.509)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.074	0.042	0.031 (0.107)
Railway Contact	0.237	0.107	0.129 (0.081)
Explorer Contact	0.143	0.286	-0.143** (0.069)
Initial Population Density / 100	11.803	13.760	-1.958 (8.008)
Distance to 1400 City (100km)	8.872	6.969	1.903* (0.932)
Distance to 1800 City (100km)	15.587	10.704	4.883** (1.890)
Observations	651	28	679

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$  The table reports summary statistics. Column (1) reports average values for missions without a printing press. Column (2) reports average values for missions with a printing press. Column (3) reports the difference between columns (1) and (2) and the significance of a standard t-test on mean differences. Standard errors are in parenthesis.



## 2.4 Empirical Analysis

### 2.4.1 Specification

Equation 2.1 describes our preferred specification

$$y_{rt} = \mathbf{Dis}'_r \beta + \mathbf{X}'_{rt} \gamma_1 + \mathbf{W}'_r \gamma_2 + \zeta_r + \eta_t + \theta_c + \varepsilon_{rt}. \quad (2.1)$$

The outcome variable  $y_{rt}$  will typically be the average prevalence rate of HIV as estimated from the DHS biomarkers, in the DHS cluster  $r$ , and year  $t$ . In the remainder of the paper, we will call “towns” the DHS clusters.  $\mathbf{Dis}$  is a vector with 4 distances from town  $r$  to the closest historical mission settlement, the closest missions with a health investment (a hospital or a dispensary), with a printing press, and with an education investment (a high or a boarding school).

The vector  $\mathbf{X}_{rt}$  contains town-level controls from each round of the DHS. These are the share of women in town, of married couples, of houses with electricity, households owning a bike or a car, the average age, and population density.

The vector  $\mathbf{W}_r$  contains geographical and historical-level controls for town  $r$  in country  $c$ . The geographical controls are distance to the capital city, distance to the coast, the average malaria ecology, ruggedness, elevation, number of growing days, and suitability for rainfed crops in the town.<sup>6</sup> The historical controls are distances to cities in 1400 and 1800, estimates of precolonial population density from HYDE, distance to colonial railways and to initial explorer routes (Nunn, 2008). Following Oster (2012), we include distance to the originating point of the epidemic as a control in all our regressions. The exact latitude and longitude of her analysis are not specified in her paper, so we use distance to Kinshasa, as it has been documented to be the most likely originating point of the HIV-1 virus (Faria et al., 2014).

Finally,  $\zeta$ ,  $\eta$ , and  $\theta$  are town, year, and country fixed effects. All the regressions’ standard errors are clustered at the closest mission level.

---

<sup>6</sup>These variables are constructed from GAEZ raster data. The constructions is explained with more details in the online Appendix

**Table 2.3** – Missions and HIV

	All	Mission at 200km		
	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Distance to mission	-0.016*** (0.004)	-0.014*** (0.005)	-0.003 (0.003)	-0.004* (0.003)
Distance to catholic	-0.004 (0.006)	-0.005 (0.008)	-0.009** (0.004)	-0.009*** (0.003)
Observations	13,216	7,545	7,545	7,482
Country and Wave FE	No	No	Yes	Yes
Historical and Geo Controls	No	No	No	Yes
Clusters	215	207	207	207
R-sq	0.17	0.14	0.42	0.44

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is the town. The dependent variable is the average prevalence rate of HIV in town. Standard errors in parentheses are clustered at the closest mission level. Controls are described in the text.

## 2.4.2 Results

**Christian Missions and HIV Prevalence** We first estimate the long-term effect of proximity to a mission station on the average HIV prevalence today. The results are reported in Table 2.3. The baseline correlation between HIV prevalence and proximity to a mission is negative, whether the mission is Catholic or Protestant. This means that regions close to historical mission settlements are more likely to have a high HIV prevalence today. As we restrict the analysis to regions close to Protestant missions and we add the controls and fixed effects described in equation 2.1, the correlation between regions close to Protestant missions and HIV prevalence vanishes and loses statistical significance. The coefficient on distance to Catholic missions becomes statistically significant and is also negative. The results indicate that a one standard deviation increase in the distance to a Protestant mission or to a Catholic mission increases the HIV prevalence rate by around 5% of a standard deviation in both cases.

As detailed in the introduction, Christian missions have mostly been associated with positive long-term effects on economic development. Therefore, the rest of the article examines the possible channels explaining the counterintuitive result from table 2.3, according to which proximity to Christian missions is associated with higher HIV prevalence.

**Missionary Investments and HIV Prevalence** In Table 2.4, we investigate the long-term effect of missionary investments on HIV prevalence today. Controlling for distance to

investments and restricting the analysis to missions close to protestant missions eliminates the baseline correlations of table 2.3 between distance to any mission and HIV prevalence.

Column (1) in table 2.4 shows that distance to missionary investments has conflicting effects on HIV prevalence. On the one hand, a one standard-deviation increase in distance to the printing press increases HIV prevalence by around 7.9% of a standard deviation. On the other hand, a one standard-deviation increase in distance to a missionary health facility decreases HIV prevalence by around 7.3% of a standard deviation. Finally, distance to education has no statistically significant effect on HIV prevalence today.

Proximity to health investments likely captures the supply of medical services today, as medical care and health care on the field is still largely supplied or coordinated by Christian institutions today. The observed relation is then likely driven by the better coverage of health care in regions close to historical mission settlements. In the section specifically addressing endogeneity, we make sure that the effect is not driven by selection on the location of health investments within the missions.

Cage and Rueda (2016) point out the long-term consequences of the printing press for outcomes that are positively associated with development, such as media consumption and social capital. How can we make sense of the negative long-term relationship between the printing press and HIV prevalence?

Duflo et al. (2015) analyze the complex relationship between education costs, sex-education programs, and HIV prevalence. They show that when no information on how and why to use condoms is taught, education has little preventive power on HIV contamination. Moreover, since most HIV prevention programs in sub-Saharan African schools are designed or coordinated by local Christian actors, it is not rare that their content emphasizes on abstinence and having a single partner, without teaching anything about condoms at all. Since the printing press was an important tool of conversion to Christianity, we test whether the negative relationship observed is linked to Christian-specific catalysts to the spread of the epidemic.

In column (2), we control for media consumption (newspaper readership) and literacy, as these are important tools for information diffusion. Consistently with Duflo et al. (2015), we find that access the means of information in itself does not change HIV prevalence. Halperin

and Epstein (2007) identify two important determinants of the HIV epidemic: the number of sexual partners (increases contagion risk) and male circumcision (decreases contagion risk). Both of them can be correlated with Christianity, as neither circumcision, multiple partners before marriage, or polygamy are encouraged in the Christian tradition. Column (3) reports the results when male circumcision and number of partners are added as controls (these variables are not present in all the rounds of the DHS, hence the decrease in the number of observations). We do observe a positive correlation between the number of partners (resp. share of male who are circumcised) and HIV prevalence. Adding these controls reduces the statistical significance of the coefficient associated with distance to the printing press, but the magnitude remains unchanged. The coefficient on distance to catholic missions remains statistically significant.

We further investigate whether the observed negative relationship between proximity to the printing press and HIV prevalence can be explained by attitudes and behaviors regarding condom use among Christian populations. We thus look at respondents' reported religion, attitudes, and beliefs about condom use and HIV that are consistently reported in the DHS. Columns (4) and (5) show the results of regression 2.1 when controlling for the share of individuals who are Christian, those who disapprove of contraception use ("against contraception"), and the share of those who think that condoms do not protect against HIV ("condoms do not protect").<sup>7</sup> The coefficients associated to the aforementioned variables all have the expected signs. Adding these variable to the regression makes the effect of distance to the printing press almost nil and insignificant, while the one of distance to a health facility is now greater and more statistically significant. Results in column (4) suggest that within regions located close to missions, a one standard-deviation increase in distance to a historical health facility increases HIV prevalence by about 11.3% of a standard deviation.

Most of the controls added in the regression reported from columns (2) to (5) do not have statistically significant effects on HIV prevalence. We thus run an F-test on the joint statistical significance of these variables. The obtained F-statistic is 2.29 (p-value=0.02), so we reject the null hypothesis of joint insignificance at the 5% level.

Because we do not have information on Catholic investments, we check that the results

---

<sup>7</sup>The variables and procedures used in the construction of all the additional controls are carefully explained in the online Appendix

are not driven by regions close to them. In column (5), we run the same regression than in column (4), but restricting it to a sample of towns close to Protestant missions (less than 200 km, as in the other regressions), but far to Catholic (more than 200 km). The results are robust to this restriction.

**Matching** Following Cage and Rueda (2016), we develop a matching strategy to tackle the possibility of an endogeneous selection of missions into investments. The previous section shows that controlling for Christian attitudes towards contraception and condom use, investments in health have decrease HIV/AIDS prevalence. Using these controls, proximity to the printing press is no longer associated to any significant effect on HIV/AIDS prevalence. Our matching strategy aims at strengthen the evidence that proximity to a historical health investment is not capturing a confounding determinant of HIV prevalence that is not related to health provision or health-seeking behaviors.

At the mission-level, we regress a binary variable indicating whether missions are endowed with a printing press or a health investment on all the observable characteristics available at the mission level (these observables correspond to the variables reported in Tables 2.1 and 2.2). From this regression we compute the propensity score, which is the estimated probability of having a printing press. We then match each mission with a health investment or a printing press to the mission with the closest propensity score using a one-to-one matching. The sample of missions with an investment is called  $In$ , where  $In$  is either a printing press ( $PP$ ), or a health investment ( $H$ ).

Each town in the DHS is then associated to the closest mission in the sample  $\{\overline{In} \cup In\}$ . We then construct the following variables:

- $D\{\overline{In} \cup In\}_r$ : the distance from town  $r$  to the closest mission station in the sample  $\{\overline{In} \cup In\}$  (“Distance to investment or likely”).
- “Treat Investment” $_r$ : a binary variable equal to one if town  $r$  is treated by one of the investments (i.e. the closest mission in the sample  $D\{\overline{In} \cup In\}_r$  actually has the investment, meaning that it belongs to  $In$ ).

Equation (2.2) describes the identification:

**Table 2.4** – Missionary Investments and HIV

	(1)	(2)	(3)	(4)	(5)
	b/se	b/se	b/se	b/se	b/se
Distance to mission	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)	-0.003 (0.002)	-0.008* (0.004)
Distance to catholic	-0.007** (0.003)	-0.007** (0.003)	-0.007* (0.004)	-0.001 (0.004)	-0.011 (0.010)
Distance to printing	-0.009** (0.004)	-0.008* (0.004)	-0.004 (0.004)	-0.003 (0.004)	0.005 (0.007)
Distance to education	-0.006 (0.004)	-0.006 (0.004)	-0.004 (0.003)	-0.005* (0.003)	0.002 (0.006)
Distance to health	0.008** (0.004)	0.007** (0.004)	0.010*** (0.003)	0.011*** (0.003)	0.015*** (0.005)
Reads news weekly		-0.019 (0.017)	-0.026 (0.021)	-0.026 (0.022)	-0.016 (0.019)
High literacy		0.044*** (0.014)	0.038*** (0.013)	0.037*** (0.013)	0.035 (0.022)
Male circumcision			-0.019 (0.019)	0.008 (0.014)	-0.026 (0.020)
Number of sex partners			0.019* (0.011)	0.010 (0.010)	0.016 (0.015)
Age gap in couples			-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Against Contraception				0.013 (0.013)	0.001 (0.015)
Condoms do not protect				0.022 (0.019)	0.004 (0.022)
Christian				-0.006 (0.012)	0.022 (0.030)
Observations	7,482	7,482	4,524	4,172	2,054
Country and Wave FE	Yes	Yes	Yes	Yes	Yes
Historical and Geo Controls	Yes	Yes	Yes	Yes	Yes
Clusters	207	207	192	172	88
R-sq	0.44	0.44	0.47	0.49	0.51

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is the town. The dependent variable is the average prevalence rate of HIV in town. Standard errors in parentheses are clustered at the closest mission level. Controls are described in the text. All regressions restrict the analysis to regions at least 200 km close to missions. Column (5) focuses on regions that are far from Catholic missions (more than 200km)

$$\begin{aligned}
y_{rt} = & \lambda_1 D\{\overline{PP} \cup PP\}_j + \lambda_2 \text{Treat } PP_j + \lambda_3 D\{\overline{PP} \cup PP\}_x \text{ Treat } PP \\
& + \zeta_1 D\{\overline{H} \cup H\}_j + \zeta_2 \text{Treat health}_j + \zeta_3 D\{\overline{H} \cup H\}_x \text{ Treat health} \\
& + \mathbf{X}'_{rt} \gamma_1 + \mathbf{W}'_r \gamma_2 + \mathbf{Chris}'_{rt} \gamma_3 \\
& + \zeta_r + \eta_t + \theta_c + \nu_{rt}
\end{aligned} \tag{2.2}$$

where the sets of controls  $\mathbf{X}'_{rt}$ ,  $\mathbf{W}_r$  are the same as before and  $y_{rt}$  is the HIV/AIDS prevalence rate. The vector  $\mathbf{Chris}_{rt}$  contains all the controls from column (5) in table 2.4, these are the Christian attitudes towards contraception and condom use, as well as the average number of partners and share of males who are circumcised.

Table 2.5 presents the results of the estimation. Our hypothesis is that controlling for Christian attitudes towards contraception and condom use, only investments in health have decrease HIV/AIDS prevalence. The hypothesis is confirmed in the analysis. Only the variables capturing proximity to a mission with a health investment is associated to lower prevalence of HIV/AIDS. The estimates suggest that regions close to a historical health investment rather than to a similar mission without a health investment (Treat  $H=1$ ), decreases prevalence rates by 0.063 points.

## 2.5 Robustness checks and Discussion

The results from the previous section suggest that missionary investments had opposing effects on HIV prevalence today. This section investigates the robustness of these results.

**Road Density** As proved by Oster (2012), proximity to trade routes, and in particular to roads is a strong predictor of HIV incidence at the cross-country level. The aim of this robustness check is twofold. First, we want to test the accuracy of our HIV prevalence measures by reproducing Oster's result. Second, we test whether our results are robust to the inclusion of road density as a control. HIV/AIDS is mostly an epidemic of urbanized areas, so although we control for light density in all our specifications, this robustness check aims at strengthening the evidence that among regions close to missions, distance to a health

**Table 2.5** – Missionary Investments and HIV, Matching Approach

	(1)	(2)
	b/se	b/se
<b>Printing</b>		
Dis pp or likely	-0.011** (0.005)	-0.004 (0.005)
Treat pp	0.030 (0.051)	0.016 (0.051)
Dis pp or likely x Treat pp	-0.003 (0.004)	-0.001 (0.005)
<b>Health</b>		
Dis health or likely	0.006 (0.004)	0.009** (0.004)
Treat health	-0.065** (0.032)	-0.063** (0.031)
Dis health or likely x Treat health	0.006* (0.003)	0.006** (0.003)
Observations	11,449	6,671
Christianity and Sex Behaviors	No	Yes
Country and Wave FE	Yes	Yes
Historical and Geo Controls	Yes	Yes
Clusters	214.00	182.00
R-sq	0.46	0.51

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is the town. The dependent variable is the average prevalence rate of HIV in town. Standard errors in parentheses are clustered at the closest mission level. Controls are described in the text. All regressions restrict the analysis to regions at least 200 km close to missions. Column (2) focuses on regions that are far from Catholic missions (more than 200km)



investment is not capturing something else linked to urbanization that would also affect HIV/AIDS prevalence rates.

To construct road density at the town level, we use data from the NASA’s Socio-Economic Data and Application center. They provide a geographic dataset with all the roads reported from country’s data and other data sources (Socio-Economic Data and Application Center (SEDAC), 2016). We keep all the primary to tertiary roads in the sample for which the existence is not claimed to be “dubious”. We then draw a 50 kmx50 km grid on the map and compute the average length of roads in each grid. We define road density as the average length on each grid, divided by the grid’s surface.<sup>8</sup>

Columns (1) and (2) in table 2.6 report the results. Column (1) shows that there is indeed a positive relationship between road density and HIV prevalence in town. These results are reassuring regarding the quality of the HIV prevalence estimates. Column (2) reports the results of running the same regression as in column (4) of table 2.4, adding road density as a control. Our results remain unchanged, but the coefficient on road density loses its statistical significance. This does not mean that Oster’s results are invalidated. Adding all the controls on distance to missionary investments means that we are comparing similar, urbanized, and more developed areas, in which the presence of a road network might make no difference for the risk of infection.

**Malaria Incidence** We also investigate the extent to which proximity to historical mission locations and investments affect malaria. Malaria is a particularly interesting counterfactual in our framework. While malaria – contrary to HIV/AIDS – is not a sexually transmitted infection, it is one of the most significant diseases in the world today in terms of its humanitarian burden. Depetris-Chauvin and Weil (2013), studying the long-term effects of malaria on economic development in Africa, find that in the more afflicted regions, malaria lowered the probability of surviving to adulthood by about ten percentage points (on the impact of malaria on growth and development see also: Cutler et al., 2010; Weil, 2010).<sup>9</sup> Malaria is a leading cause of death for children and the cause of numerous lost work hours for adults nowadays (Cohen et al., 2015).

---

<sup>8</sup>The normalization matters because the cell’s size is smaller when cut by a coast

<sup>9</sup>On the long-term effects of other diseases in Africa, see Alsan (2015).

Malaria can be prevented in a number of cheap ways, in particular different class of antimalarials, among which the artemisinin combination therapies (ACTs), the provision of insecticide-treated nets (ITNs) and indoor residual spraying (IRS, meaning the spraying of insecticide in house's walls). A natural way to ensure access for appropriate users while limiting over-treatment is to distribute subsidized health technologies like ACTs, IRS, and ITNs, through the public health system, where diagnostic tools and trained medical personnel can target technologies to patients with high returns (Cohen and Dupas, 2010; Cohen et al., 2015). Cohen and Dupas (2010) show the effectiveness of public policy intervention to incentivize ITN use and the importance of peer-effects that encourage uptake.

In table 2.6, we present the results of estimating equation 2.1 using malaria incidence in 2015 as an outcome. The estimates of malaria incidence come from the maps produced by the Malaria Atlas Project (Bhatt et al., 2015). Malaria prevalence pre-existed missionary location. On the one hand, missionaries may have avoided malaria to protect their own health, but on the other hand, in the quest of "healing" the most deprived populations, it may as well have positively affected it. In particular, table 2.1 shows that among missions, those with a health investment tend to have higher malaria ecology. Moreover, the mapping of malaria incidence has greatly changed over time, with the expanding use of technologies to prevent contagion (Bhatt et al., 2015). For the aforementioned reasons, we do not seek to interpret the coefficient on missionary investments and malaria incidence today.

However, given that malaria is not sexually transmissible, malaria should not be affected by Christian influence on sexual behaviors per se. Therefore, the inclusion of the controls on sexual behaviors and beliefs that are linked to Christianity should not affect the observed relationships, as opposed to what is observed in table 2.4 for HIV prevalence. Columns (3) and (4) in table 2.4 show the results. The inclusion of the controls proxying for Christian influence on sexual behaviors does not change the correlation between distance to investments and malaria incidence in 2015. An F-test on the joint-significance of the added controls returns a F-statistic of 1.01 ( $p$ -value=0.42). As opposed to the results for HIV/AIDS prevalence, we thus fail to reject the null that the controls for Christian beliefs on contraception and condom use as well as male circumcision and number of partners do not affect the outcome. It is therefore unlikely that this set of controls captures something else directly related to

health prevention in general.

## 2.6 Conclusion

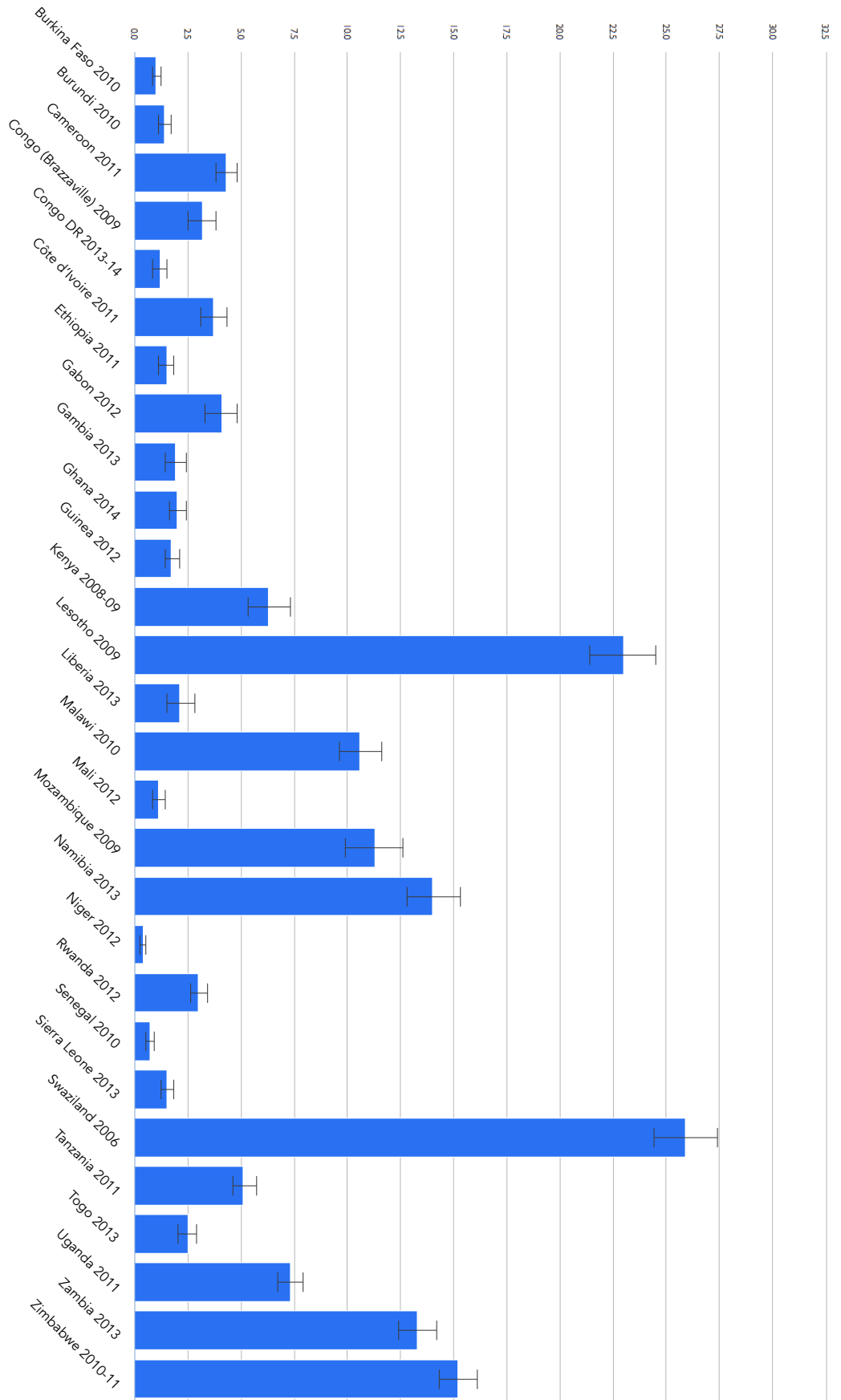
In this paper, we provide new empirical evidence on the long-term effects of missionary activities in sub-Saharan Africa on the prevalence of HIV/AIDS nowadays. We highlight two possible countervailing effects of missions. On the one hand, early missionary investments in health have a positive long-term impact on prevalence as of today. On the other hand, the impact of missionary activity is also negative effect through its effect on behaviors regarding sexual health and safety. By focusing on moral values and abstinence before marriage, and ignoring condoms, they increase the risk of HIV/AIDS within comparable regions.

While the focus of the existing literature have being mostly on missionary presence per se, to the exception of Cage and Rueda (2016), we highlight the importance of distinguishing the investments performed by the mission to the missionary presence and in particular the Christian influence. Through their investments in health, missions have a positive long-term effect on development, consistently with the existing literature, this positive effect is counter-balanced by the negative Christian influence on beliefs and the practice of safe sex methods. These results also have important policy implications in terms of optimal policy to reduce the incidence rates in the future.

**Table 2.6** – Robustness Checks

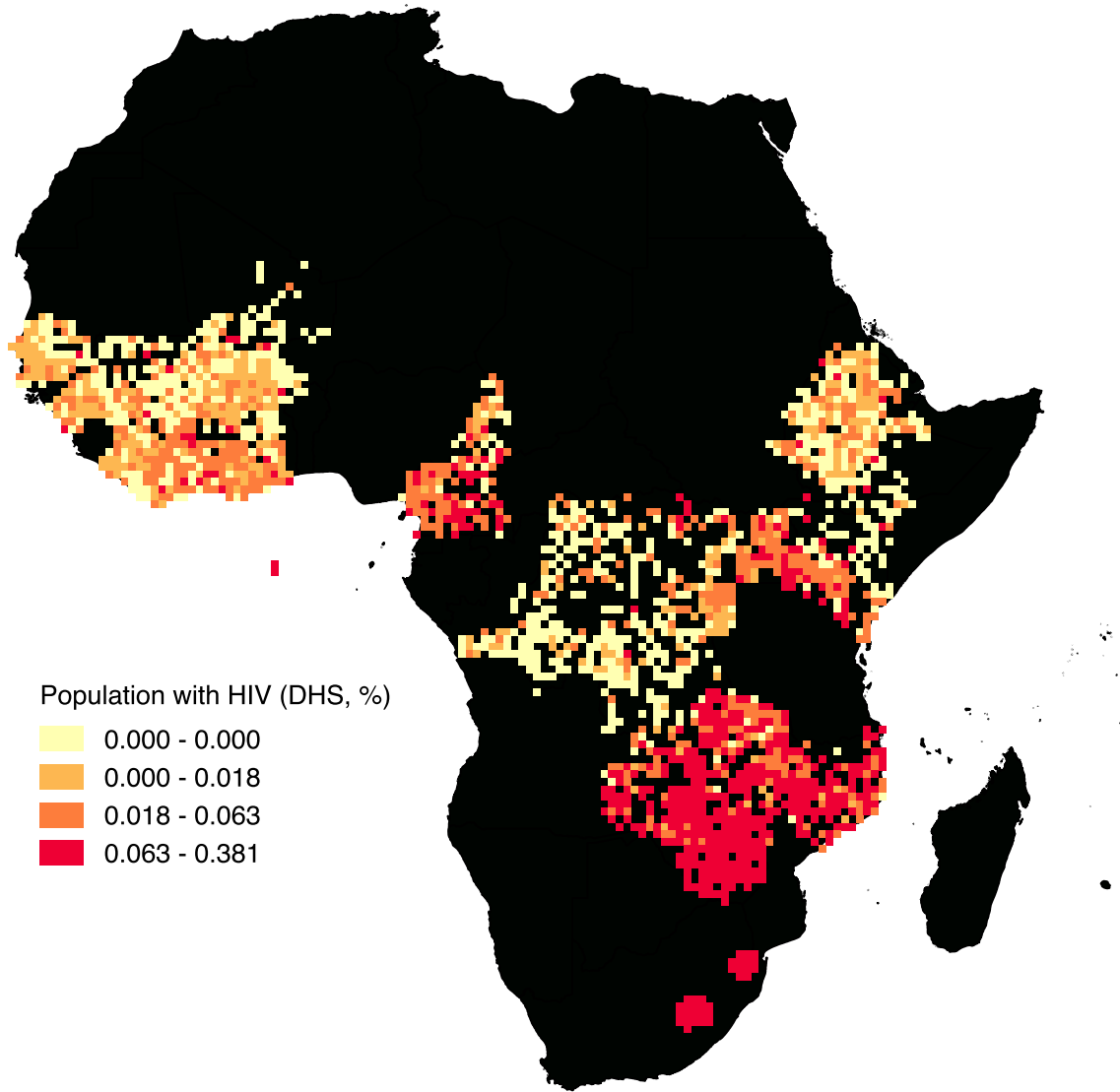
	HIV Prevalence		Malaria Incidence	
	(1) b/se	(2) b/se	(3) b/se	(4) b/se
Road Length per km2	0.140** (0.058)	-0.006 (0.043)	0.161** (0.077)	0.151* (0.083)
Distance to mission		-0.001 (0.003)	0.024*** (0.005)	0.025*** (0.005)
Distance to catholic		-0.007 (0.004)	-0.012** (0.006)	-0.013** (0.006)
Distance to printing		-0.004 (0.004)	0.019*** (0.007)	0.019** (0.007)
Distance to education		-0.004 (0.003)	0.006 (0.006)	0.007 (0.006)
Distance to health		0.009*** (0.003)	-0.020*** (0.007)	-0.021*** (0.007)
Number of sex partners		0.016 (0.010)		0.006 (0.010)
Age gap in couples		-0.001 (0.001)		0.001 (0.001)
Against Contraception		0.011 (0.011)		-0.009 (0.009)
Condoms do not protect		0.037** (0.019)		-0.020* (0.011)
Christian		-0.007 (0.011)		0.019 (0.013)
Observations	7,482	4,524	7,482	7,102
Country and Wave FE	Yes	Yes	Yes	Yes
Historical and Geo Controls	Yes	Yes	Yes	Yes
Clusters	207	192	207	207
R-sq	0.40	0.47	0.77	0.76

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is the town. The dependent variable is the average prevalence rate of HIV in town (columns (1) to (4)), and the average incidence rate of malaria in town in 2015 (columns (5) and (6)). All regressions restrict the analysis to regions at least 200 km close to missions. Standard errors in parentheses are clustered at the closest mission level. Controls are described in the text.



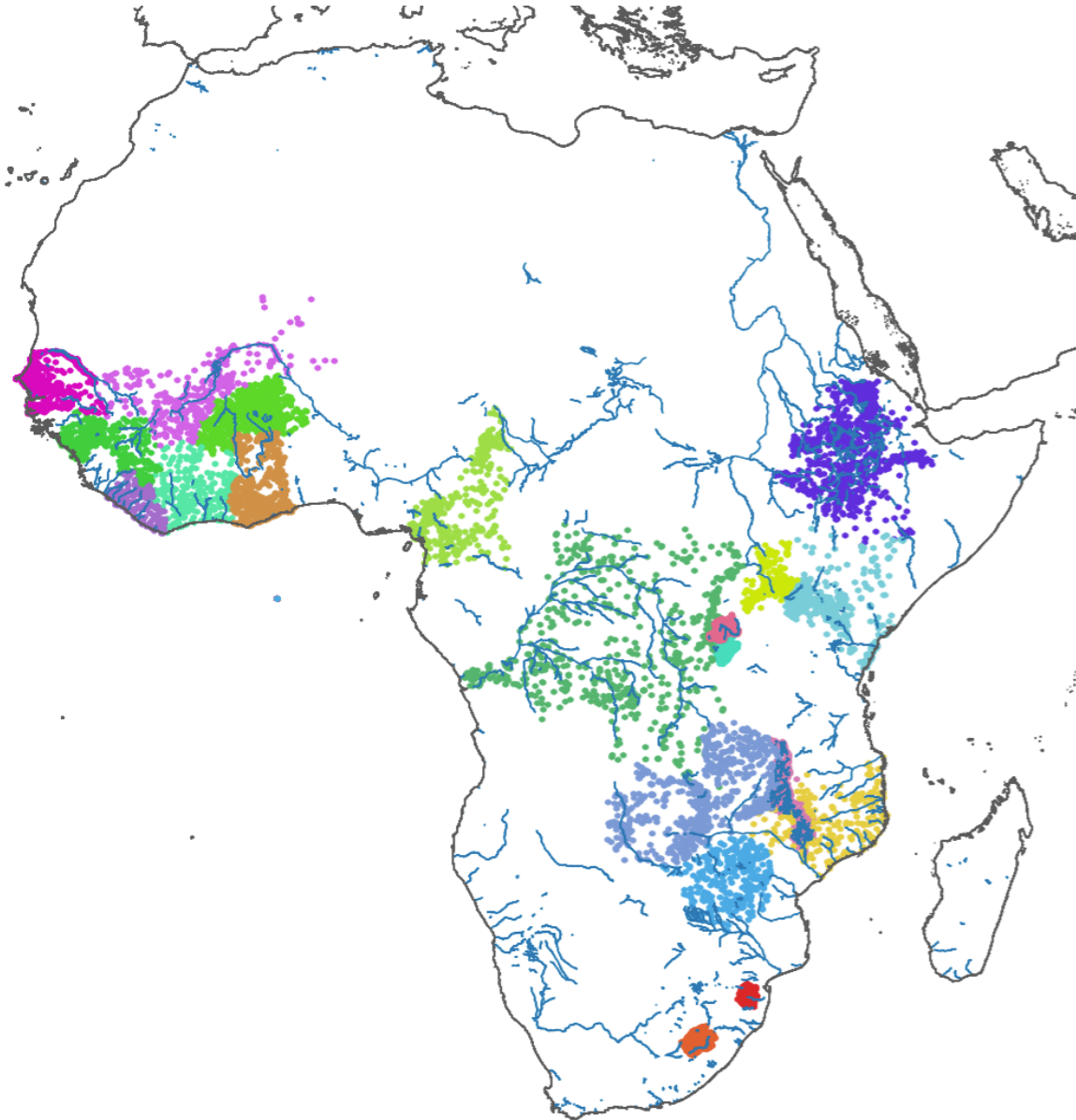
**Figure 2.2** – HIV prevalence rates by country

**Notes:** The figure shows the average prevalence rates from different sub-Saharan African countries. The estimates are those of the DHS using the most recent biomarker data for each country (extracted from statcompiler).



**Figure 2.3** – HIV prevalence estimates at the subnational level

**Notes:** The figure shows the average prevalence rates from DHS clusters, estimated by the authors using available biomarker data from 2003 to 2013. The figure shows the average estimated prevalence from all the DHS cluster in a 50x50 km grid.



**Figure 2.4** – DHS locations

**Notes:** The figure shows the location of DHS clusters in sub-Saharan Africa that report both HIV biomarker data and GPS information. Each color represents a different country.

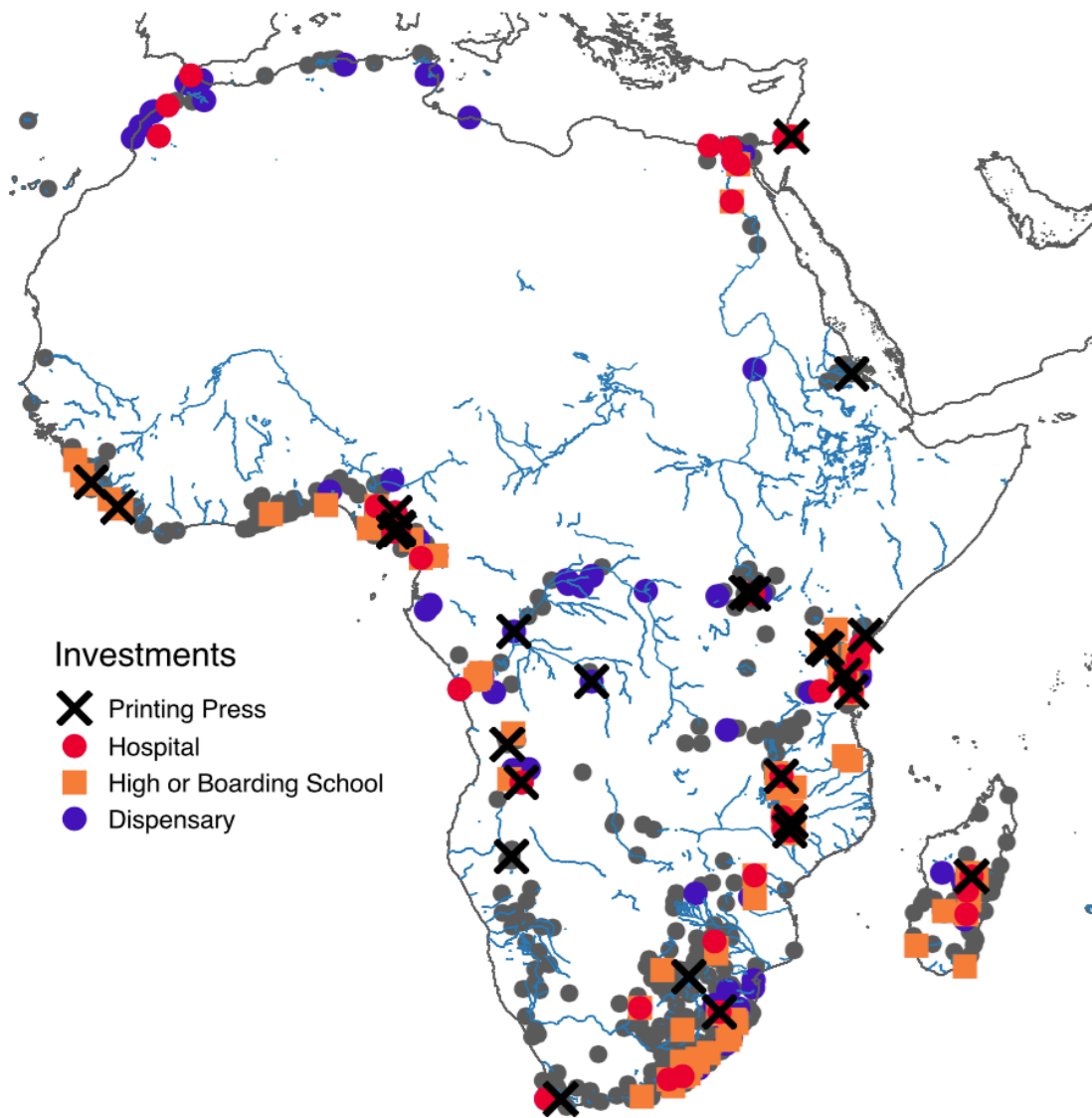
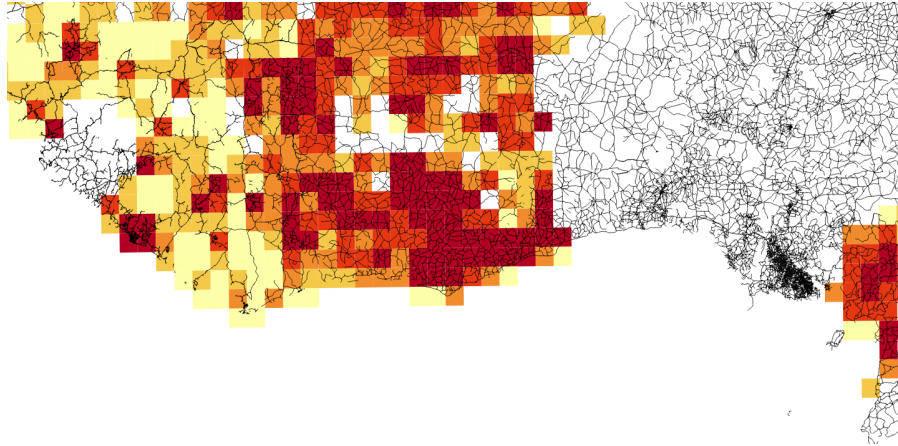


Figure 2.5 – Missions and missionary investment locations

Notes: The figure shows the location of missions and our investment of interest: health investment, schools, and the printing press





**Figure 2.6** – Road Density

**Notes:** The figure a part of the map of African roads from the SEDAC and the extracted road density measure at the grid level (in Western Africa). The colors indicate a range of road densities from low (light color) to high (dark color). The areas that are non gridded correspond to regions where no DHS town is located.

## Part II

The role of ancestry to explain  
individual economic success at the  
local level in the United States



## Chapter 3

# What Population Genetics can and cannot tell us about societies

This chapter introduces and explains the notions of population genetics used in Economics. I have relied on a variety of sources, in particular on Cavalli-Sforza et al. (1994) and Relethford (2012), to articulate my understanding of these notions and of their interpretation in the social sciences. The presentation rely sometimes on simplifications, and all the errors remain my own.

In my research, I use bilateral measures of genetic distance that account for the extent of population exchanges (mating) between groups. For the sake of simplicity, I call this first set of measures the “between group measures”. This approach is inspired by the research of Spolaore and Wacziarg (2009); Desmet et al. (2015); and Spolaore and Wacziarg (2015). There is another strand of the literature that uses genetic diversity at the population level, as opposed to measures aimed at doing comparisons *between* . I call this second set of measures the “within group measures”. This strand of the literature is exemplified by Ashraf and Galor (2013) who claim that genetic diversity has influenced economic development over hundreds of thousands years.

The aim of this chapter is not to present original empirical results. It is to be read instead a toolbox for a critical understanding of the Economics literature using genetic data. The first section presents the “within group measures”. The second section presents the “between group” measures and how they can capture cultural divergences between populations. The

third section concludes.

### 3.1 Within-population measures: using genetic diversity in the Social Sciences

An increasing number of research in Economics and in the social science has been using genetic diversity, heterozygosity, as a predictor for economic success. The most cited piece of research using such measure is Ashraf and Galor (2013) (that I will now refer to as AG), which finds an inverse-U shaped relationship between expected heterozygosity and economic development.

They interpret genetic diversity as a form of diversity in population that “confers both social costs, in the form of miscoordination and distrust arising from genetic differences across the members of society, and social benefits in the form of diversity-driven knowledge accumulation.” Although the wording is careful, this hypothesis leaves room for an interpretation of heterozygosity as a proxy for diversity in preferences, cultures, or skills, that could explain coordination problems or greater creativity.<sup>1</sup>

What does heterozygosity exactly measures? How is it expected to evolve over time? Why does it decrease with migratory distance from East-Africa? Can we interpret it as a proxy for general diversity in societies?

This section will present the biological definition of heterozygosity and its evolution over time. Using these definitions, I will then discuss the interpretation it can bear in Economics.

---

<sup>1</sup>The ambiguity in the interpretation is also evident in a response that Ashraf and Galor addressed to their critics:

“The key is that the measure of intra-population genetic diversity that we employ should be interpreted as a proxy (i.e., a correlated summary measure) for diversity amongst individuals in a myriad of observable and unobservable personal traits that may be physiological, behavioral, socially-constructed”.

This reply was published on Brown’s department of Economics website and is no longer online. Excerpts can be found on Jason Collin’s blog: [jasoncollins.org](http://jasoncollins.org)

### 3.1.1 Heterozygosity

The measure of genetic diversity used in AG is heterozygosity. Heterozygosity is the probability that two particular loci taken at random from the population have a different allele.

An allele is defined as one of the possible forms a gene or a piece of chromosome can have (a locus). Consider a population with a number  $N_A$  of different alleles for a given locus. Each allele  $a$  has frequency  $p_a$  in the population. Then heterozygosity is defined as

$$H = 1 - \sum_{a=1}^{N_A} p_a^2 \quad (3.1)$$

The frequency of alleles in a population can vary for many reasons. This is particularly the case if there is natural selection pressure on some alleles. A textbook example of this is the HBB gene on human chromosome 11 that encodes beta haemoglobin. One possible allele of this gene is called the *S allele*. This allele is responsible for sickle-cell disease, which in turn is known to protect against certain forms of malaria. Because the *S allele* cause sickle-cell disease, which is a serious condition, its proportion is usually close to 0. However, in regions where there is high malaria prevalence the proportion can reach around 12% (Relethford, 2012, p.184). Such difference in proportions is then due to different environmental pressures.

For this reason, the measures of heterozygosity in the social sciences have concentrated in *non-coding regions of DNA* that do not have known functions to affect our phenotypes. In that case, heterozygosity is less likely to capture environmental or geographical pressures. In other words, computed with non-coding regions of DNA, heterozygosity measures capture allele diversity that is the result of random genetic drift and initial conditions. What does this diversity capture for population histories?

### 3.1.2 Heterozygosity and Genetic Drift

Consider a simplified population with  $n$  individuals that works as follows. Each individual has two sets of chromosomes, one for each parent. There are thus a total of  $2n$  chromosomes. A second generation population of the exact same size then derives from this initial population with random and non-overlapping mating. This simplification then models the second generation as a random draw with replacement of  $2n$  loci from the first generation.

This implies that for each locus in the second generation, the probability of being a copy of a particular locus in the first generation is exactly equal to  $\frac{1}{2n}$ . Referring to figure 3.1, the probability that any  $g_i^2$  is a copy of any particular  $g_j^1$  is identical for all  $i$  and  $j$  and equal to  $\frac{1}{2n}$ .

We say that for any two loci in generation  $t$ , they are identical by descent (IBD) if and only if they are identical because they derive from the same ancestor. In a simplified population like the one we are modeling here, which has no mutation, the probability of identity by descent for two loci is thus  $1 - H$ .

Denote  $f_t$  the probability that two loci are IBD. By construction,  $H_t = 1 - f_t$ , where  $H_t$  is heterozygosity in generation  $t$ . We will now derive the probability  $f_t$ . Gene copies have a common ancestor either because they have the same “parent” from generation  $t - 1$  (case 1) or because they had a common ancestor before  $t - 1$  (case 2). It follows that

$$f_t = P(\text{case 1}) + P(\text{case 2}).$$

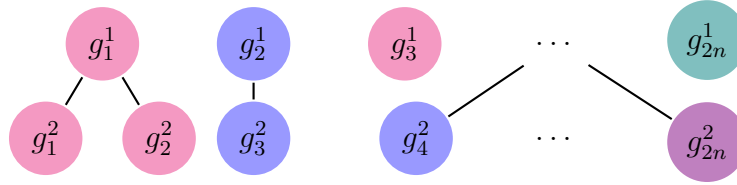
The probability of case 1 is the probability that any pair of loci from generation  $t$  are copies of the same locus in generation  $t - 1$ . For any pair, the probability of this is  $\frac{1}{4n^2}$ . Since there are  $2n$  possible original loci in generation  $t - 1$ , then

$$P(\text{case 1}) = \frac{1}{2n}.$$

The probability of case 2 is the probability that any pair of the loci from generation  $t$ , does not descend from the same locus in generation  $t - 1$  (probability= $1 - \frac{1}{2n}$ ) AND that they have a common ancestor from before  $t - 1$  (probability= $f_{t-1}$ ). It then follows that:

$$P(\text{case 2}) = f_{t-1} \left(1 - \frac{1}{2n}\right).$$

**Figure 3.1** – An illustration of common descent



**Notes:** In this representation alleles  $g_n^t$  are the alleles number  $n$  from generation  $t$ . The different colors represent the different alleles. In this diagram,  $g_1^2$  and  $g_2^2$  have common descent from generation 1 because they both come from  $g_1^1$ . On the contrary,  $g_4^2$  and  $g_3^2$  do not have common descent from the previous generation. However, because they have the same allele (blue), they can still share a common ancestor from a previous generation

From the previous results, we can derive:

$$\begin{aligned}
 f_t &= \frac{1}{2n} + f_{t-1} \left(1 - \frac{1}{2n}\right) \\
 \Leftrightarrow \underbrace{1 - f_t}_{H_t} &= (1 - f_{t-1}) \left(1 - \frac{1}{2n}\right) \\
 \Leftrightarrow H_t &= H_0 \left(1 - \frac{1}{2n}\right)^t
 \end{aligned} \tag{3.2}$$

Equation 3.2 shows that in this simplified population, heterozygosity will tend to decrease at rate  $\frac{1}{2n}$ . So the smaller the initial population size, the faster will heterozygosity decrease. This result sheds light on the “out of Africa” assumption. We expect that heterozygosity of the populations that left Africa tens of thousands years ago will be smaller because they descend from a subsequent subsamples of the “original” population (which will already restrict the pool of genes), and because as a result of the decrease in population size (due to subsequent settlement), heterozygosity is also expected to decrease faster.

### 3.1.3 Heterozygosity in the Social Sciences

The meaning of heterozygosity for Economics is ambiguously defined in AG. The language of AG focuses on establishing a precise causal relationship: “the level of genetic diversity within a society is found to have a hump-shaped effect on development outcomes” (Ashraf and Galor, 2013). There is almost no cultural or societal interpretation of their measure of diversity. The only general interpretation of genetic distance is to be found in the concluding remarks of the article, when they interpret their findings as proof that: “deep-rooted factors



determined tens of thousands years ago, had significant effect on the course of economic development from the dawn of human civilization to the contemporary era”. They posit that the hump shaped effect of the relationship is due to lowered levels of coordination and trust for highly heterozygous populations and lowered levels of innovation for low heterozygosity populations. These proposed channels channels do suggest, nevertheless, an extrapolation between genetic diversity and social or cultural diversity.

This section shows that such extrapolation is not convincing for two main reasons. First, the measures of heterozygosity used in AG are not precise enough to support the claim. This matter of internal validity is mostly a problem for African and Latin-American populations, which are the extreme values weighting for the hump-shaped relation. Second, another possible interpretation of their results links heterozygosity to population size. I show that this is not a suitable interpretation. Finally, although AG’s claims are not empirically convincing, their hypothesis and methods are a fertile terrain for research, which I will discuss in more depth.

### **Measuring heterozygosity in contemporary Africa and Latin-America**

AG generated a heated reaction among academics. A group of Harvard anthropologists, evolutionary biologists, and medical doctors even published a public objection to the article (d’Alpoim Guedes et al., 2012, 2013). One of the most interesting points in the objection is that populations with particularly low levels of diversity reached extremely high levels of urbanization before colonization, such as the Aztecs, the Incas, and the Mayas. This type of phenomena cannot be accounted in AG because the Aztec and Maya populations had already declined, and Peru and Chile are not in their sample from 1500CE. Moreover, their estimates of population for Latin American regions in 1CE and 1000CE could be “strikingly lower than most values in the archaeological literature” (d’Alpoim Guedes et al., 2013).

This point is illustrative of the complexity of interpreting AG’s relationship for the extreme values of predicted heterozygosity; the regions that push the relationship downwards to obtain the inverse-U shaped relationship. In both the Latin-American and the African cases, these are the regions that were the most affected by extractive colonization, which in turn pushes development downwards. Furthermore, these are also the regions for which

much noise is to be expected in the prediction of heterozygosity.

For African regions, distance to Addis Ababa is an irrelevant proxy for heterozygosity, as modern humans are thought to have several “originating points”, likely located in the Southern African region (Tishkoff et al., 2009). Very broadly sketched, African populations do not descend, in general, from the populations that migrated out of Africa around 60000 years ago (Laval et al., 2010). Instead, they descend from the larger pool out of which were drawn the out-of-Africa migrants; and if we are to go further in time to find their originating population, it is located somewhere far south from Addis Ababa.

For Latin America, the identification of the groups for which to compute the different migratory distances is particularly challenging. Colonization and the Atlantic trade generated drastic population exchanges, which in turn implies that if heterozygosity can remain low for some isolated populations, this claim cannot be generalized to most countries with a high degree of *mestizaje*. In some regions, the ancestry proportion of the population can be above 50% European, implying a much higher expected level of heterozygosity than the one predicted in

### **Heterozygosity and Population Size**

The previous section shows that for isolated populations with random mating, heterozygosity is mostly capturing initial population size. In that sense, we could understand AG’s claim that gene diversity can be correlated with innovation if we think that initial population size, reflected in heterozygosity, affects the probability of innovation. If disruptive innovations are rare events, having larger population pools might increase their probability at each point in time.

However, such interpretation is rather weak. Firstly, the previous section’s model, constrained population size to remain the same over generations. This needs not to be the case, since low-heterozygosity populations such as the Incas or the Mayas went to have extremely high populations for their times. The Population Genetics theory shows that in such a case heterozygosity is *still* expected to decrease faster, because the initial set of alleles is smaller. It is nevertheless no longer possible to extrapolate the notion of heterozygosity to real population size (as opposed to the number of initial alleles, which is a different concept). The

second reason is that the model focuses on isolated populations: there is no migration in the model that would lead to a substantial gene flow. This second reason is obviously the most problematic given the large populations movements observed in History, from the crusades, colonization, or the Atlantic slave trade, to the Great Atlantic Migration from Europe to the USA, or the Chinese diaspora.

### **Still a powerful hypothesis**

Although the correlation between heterozygosity and income is not as convincing as AG paper's suggest, there are interesting economic lessons we can take from it.

Firstly, although heterozygosity as used in the article is a poor measure for social or cultural diversity, this does not mean that all genetics measures are to discard for this use. In particular, the analysis of genetic variations for different populations can lead to accurate conclusions about gene flow, that is to say the amount of mating that there has been between populations. Next section explores this type of measure and their correlation to other measures of cultural variation.

Secondly, the hypothesis that diversity can have conflicting or heterogeneous effects on growth is also particularly interesting and could show why the literature on ethnic fractionalization or polarization has often disagreed on the existence of a strong correlation between ethnic diversity and conflict (Esteban et al., 2012).

Finally, although there has been a wide increase in the literature on the persistent roots of economic development, AG take one of the most long-term perspective on the question that does not rely on purely geographic factors (unlike Diamond (2010)). Is it true that diversity will affect development in such a long-term perspective? Chapter 4 of this thesis challenges this view. I show that focusing on the USA, which restricts institutional differences, diversity in ancestry is not a robust predictor of economic success. Its role depends on local social capital.

## 3.2 Between Population Measures: is genetic distance a proxy for cultural distance?

### 3.2.1 Gene flow

In the previous section, we focused on within-population measures of allelic diversity. The baseline models underlying these measures focus on isolated populations and describe their evolution resulting from random drift without considering population exchanges. However, a social scientists' intuition of diversity is probably the opposite of the account of an isolated population. The interest is exactly knowing what happens when populations stop being isolated and mix with each other.

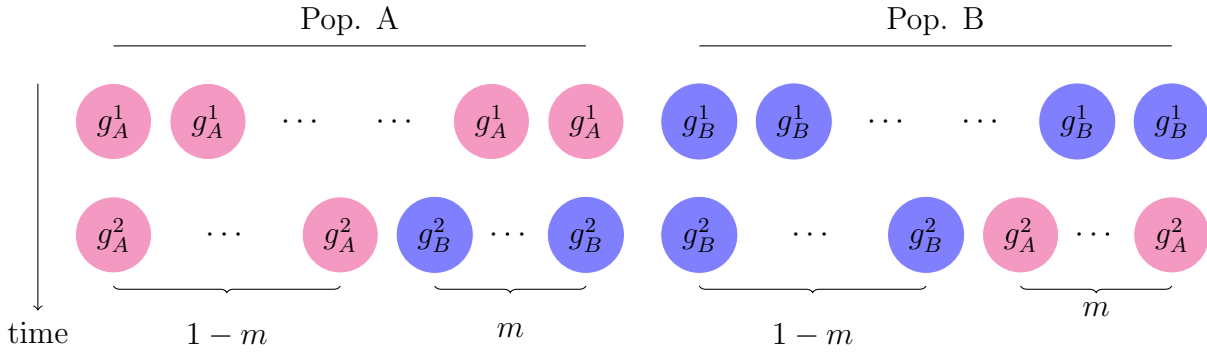
The theory of population genetics can tell us how we expect population genetics to change when we observe gene flow. Gene flow is the phenomenon of “foreign” genes mixing with a local population. It can be correlated with migration, although the two phenomena are not exactly identical. There will only be gene flow in the encounter of two separate populations if they mate with each other.

In this section, we introduce a simple model of gene flow, explaining the expected evolution of heterozygosity between two populations that enter in contact. The model is adapted from Relethford (2012, p.214).

Consider two populations,  $A$  and  $B$ , of same size and mass equal to 1 that are initially completely homozygous. The setup of this model is represented in figure 3.2. Populations  $A$  (resp.  $B$ ) has a proportion of 100% of what we call allele  $A$ , represented by  $g_A$  in pink in figure 3.2 (resp.  $B$ , , represented by  $g_B$  in light blue ). The two populations come in contact as follows: at each generation, a share  $m \in [0, 0.5[$  of alleles from  $B$  (resp.  $A$ ) gets transferred to population  $A$  (resp.  $B$ ).

Denote  $p_{At}$  the share of alleles  $A$  in population  $A$  at time  $t$ , and  $p_{Bt}$  the share of alleles  $B$  in population  $A$  at time  $t$ . Let  $P_t$  be the column vector with  $p_{At}$  and  $p_{Bt}$ . By construction of the model,

**Figure 3.2** – A model of gene flow



**Notes:** The first generation  $g^1$  starts with two initially homozygous populations. Generations after generation, they exchange a share  $m$  of their loci.

$$P_t = \underbrace{\begin{bmatrix} 1 - m & m \\ m & 1 - m \end{bmatrix}}_M P_{t-1}.$$

The transition matrix  $M$  has two eigenvalues equal to 1 and  $1 - 2m$ . Using the initial conditions of the model  $P_0 = [1, 0]'$ , we can describe the dynamics of allelic frequency in population  $A$  with equation 3.3.

$$P_t = \begin{bmatrix} 0.5 (1 + (1 - 2m)^t) \\ 0.5 (1 - (1 - 2m)^t) \end{bmatrix} \tag{3.3}$$

Using the same reasoning, we can derive  $Q_t$ , the column vector of allelic frequencies of alleles  $B$  and  $A$  (in this order) in population  $B$ , to find that  $Q_t = P_t$ . This simple model of population mixing shows that the gene flow from one population to the other makes that two separate and distinct homozygous populations will converge to be heterozygous with the same proportion of each alleles in both populations.

### 3.2.2 Genetic distance

One of the most common measures of genetic distance used in the literature is the  $F_{ST}$ . This measure aims at capturing how much of the gene variation in a group is due to between group variation as opposed to within group variation.

In section 3.1.2, we showed that random drift pushes isolated populations towards homogeneity, with the only counteracting force being mutations. It follows that on average, isolated populations will tend to diverge because genetic drift pushes the fixation of a single allele in the entire population. On the contrary, section 3.2.1 shows that gene flow between populations pushes the populations to be more alike. Therefore, studying the patterns of gene variation between and within populations is a relevant source of information to determine the likelihood that two populations have come in contact with each other in the past.

The fixation index  $F_{ST}$  is the ratio of between group genetic diversity over total genetic diversity. Consider a group of  $K$  populations each with heterozygosity  $H_k$ , and population size  $n_k$ ,  $k \in \{1, \dots, K\}$ . We denote  $\bar{H}_w$  the average of all these heterozygosities, that is the average of the heterozygosities within the populations. The total heterozygosity  $H_T$  is the expected level of heterozygosity in a theoretical population where all the loci from the  $K$  populations were pooled together. The fixation index is then defined as the ratio of the genetic variation that is not explained by within group variation over the total genetic variation. In other words:

$$F_{ST} = \frac{H_T - \bar{H}_w}{H_T} \quad (3.4)$$

This ratio varies between 0 and 1. In the extreme case where all the  $K$  populations are completely homozygous, then we expect:

$$\begin{aligned} \bar{H}_w &= 0 \\ H_T &= 1 - \sum_{k=1}^K \frac{1}{n_k^2} \end{aligned} .$$

In this particular case, then  $F_{ST} = 1$  because all the variation comes from the variation between the groups. In the other extreme case, all the subpopulations have identical allele distributions, then the variation within each group is identical to the total sample variation. In this case,  $F_{ST} = 0$ .

Why does this analysis of variance yield information on the genetic distance between groups? The model of gene flow studied in the previous section provides a useful illustration of the interest of studying gene variation to infer genetic distance between groups.

Consider again populations  $A$  and  $B$  at a time  $t$  where the share of alleles  $A$  and  $B$  in population  $A$  are respectively  $p_{At}$  and  $p_{Bt}$ . Similarly, the share of alleles  $A$  and  $B$  in population  $B$  are respectively  $q_{At}$  and  $q_{Bt}$ .

In this setup, the within-group heterozygosities of populations  $A$  and  $B$  at time  $t$  are, respectively,  $H_{wt}^A$  and  $H_{wt}^B$ :

$$\begin{aligned} H_{wt}^A &= 1 - p_{At}^2 - p_{Bt}^2 \\ H_{wt}^B &= 1 - q_{At}^2 - q_{Bt}^2 \end{aligned} \quad (3.5)$$

Using expressions from 3.3, we obtain In this setup, the average within-group heterozygosity,  $H_{wt}$  is defined by:

$$H_{wt} = 0.5 (H_{wt}^A + H_{wt}^B) = \frac{1}{2} - \frac{(1 - 2m)^{2t}}{2} \quad (3.6)$$

If we consider the two populations as a single group, then the total heterozygosity  $H_{Tt}$  is given by:

$$\begin{aligned} H_{Tt} &= 1 - r_{At}^2 - r_{Bt}^2 \\ \text{with } r_{At} &= 0.5 (p_{At} + q_{At}) \\ r_{Bt} &= 0.5 (p_{Bt} + q_{Bt}) \end{aligned} \quad (3.7)$$

In our case, because th two populations are of exact same size and start as completely homozygous, then  $r_{At} = r_{Bt} = 0.5$ . This in turn implies that  $H_{Tt} = 0.5$ . As a result, the fixation index at time  $t$  between populations  $A$  and  $B$ ,  $F_{STt}$  is given by:

$$F_{STt} = (1 - 2m)^{2t}. \quad (3.8)$$

The distance between the two populations is maximal and equal to 1 when at  $t = 0$  the two populations are isolated and with nil heterozygosity. As the population mixes, the distribution of alleles becomes gradually more similar, bringing  $F_{STt}$  towards 0.

Gene flow makes the distribution of alleles similar between the groups. This type of phenomena is what the  $F_{ST}$  will capture as a reduction in genetic differences. The more gene flow there is between groups, the faster the fixation index will decrease.

### 3.2.3 Genetic and cultural distances

The previous section shows that the genetic distance measured by the  $F_{ST}$  will capture the extent of gene flow that there has been between populations. For this reason, this measure has drawn attention as a potential proxy for cultural differences.

In population genetics, Cavalli-Sforza et al. (1994) provide a first detailed estimation of genetic distance using initially 491 populations then aggregated to 42 representative ones. Figure 3.3 shows the baseline results from this analysis. The genetic distance between populations using ranges between 0.43 for between the African San and the Australian populations to 0.021 between the Danish and the English. Using more modern genetic and statistical tools, new estimates of the  $F_{ST}$  have been published recently covering regional and worldwide analysis. Read Pemberton et al. (2013) for a detailed review.

Because genetic variation captures the extent of population exchanges, it has also been tested as a proxy for cultural similarities. Thinking of culture in a broad sense as a series of traits that determine behavior as a group level that can be transmitted over generations, it is likely that large events of gene flow exchanges can capture cultural exchanges to a certain degree.

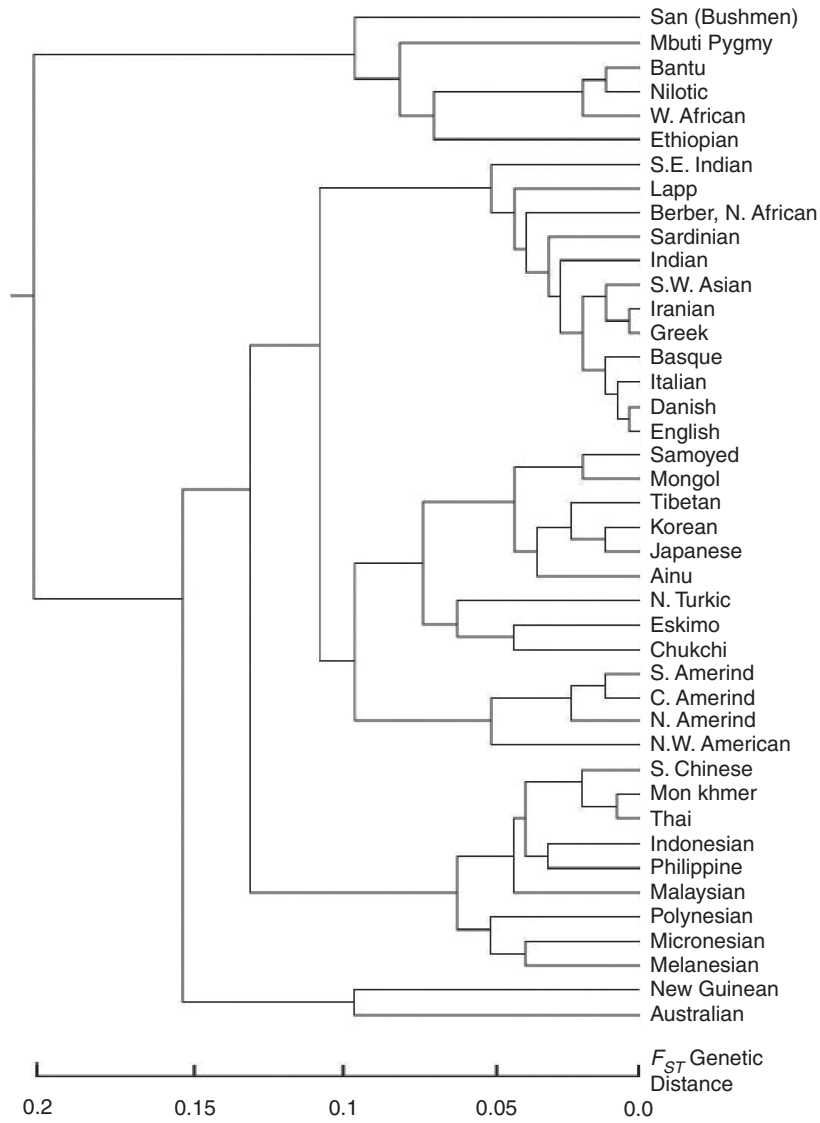
In anthropology and population genetics, there has been a long-lasting research on the correlation between genetic and linguistic distance. There is strong evidence that linguistic groups correspond to clusters of small genetic distances, as early shown by Cavalli-Sforza et al. (1988). This association has been reproduced at a regional level, finding a strong association in Europe (Barbujani and Sokal, 1990) and lesser so in Central Asia (Comas et al., 1998). A recent large-scale study among African populations confirmed the strong association between linguistic and genetic clusters of variation (Tishkoff et al., 2009).

Language is not the only cultural trait that varies with genetic distance. Brown et al. (2013) shed light on the correlation between genetic and traditional music distance. Similarly, Ross et al. (2013) show that in Europe, genetic diversity and folktales are also structured in similar clusters, although there is evidence of larger exchange in genes than in the folktales.

Using the World Value Survey (WVS), Spolaore and Wacziarg (2015) also construct a series of cultural distance and cultural diversity measures using different traits: language, religion, and values. The WVS's questions are divided into broader categories that relate to



**Figure 3.3** –  $F_{ST}$  between 42 populations, from Cavalli-Sforza (1994)



**Notes:** This phylogenetic tree shows the genetic distance between 42 populations as estimated by Cavalli-Sforza et al. (1994). The length of the branches represent the  $F_{ST}$  between the nodes.

the perceptions of life (like trust, happiness, etc.), work, family values, politics and society, and religion and morale. For each question  $i$ , they construct a measure of cultural distance for a pair of countries  $\{c_1, c_2\}$  that is given by equation 4.2

$$CD = \|s_i^{c_1} - s_i^{c_2}\| \quad (3.9)$$

Where  $\|\cdot\|$  is the Euclidian Norm and  $s_i^c$  is the vector giving the shares of people choosing each of the question's possible answers.<sup>2</sup> To create summary indexes, they standardize the question-specific distances to have zero mean and a standard deviation equal to 1. Finally, they sum the question-specific exes for all the categories in the WVS.

Using the World Value Survey, it is thus possible to correlate cultural distance from equation 4.2 and genetic distance. Using ethnic-composition by country, they obtain pairwise-measures of weighted genetic distances between countries from Cavalli-Sforza's ethnic-level genetic distance. Their analysis of the genetic  $F_{ST}$ 's predictive power for cultural differences concludes that genetic distance tends to be "*broadly and significantly correlated with a vast range of differences in cultural traits. Thus, while specific correlations with individual sets of traits are typically moderate in magnitude, there is an overall relation between ancestry and culture, consistent with a conceptual framework in which a broad range of cultural traits are transmitted with variation across generations over time*" (Spolaore and Wacziarg, 2015).

The graphs in figure 3.4 illustrate the point made by Spolaore and Wacziarg (2015). Genetic distance is robustly correlated with all of the measures, although the correlation is weaker for some dimensions (as it is the case for family values). However, genetic distance is broadly and significantly correlated with all of the dimensions. Figure 3.5 shows the correlation between the real  $F_{ST}$  values and the predicted ones in a regression of the  $F_{ST}$  on the different measures of cultural distance from the WVS. There is a strong correlation between these two variables ( $R^2 > 0.98$ ), which in turn suggests that genetic distance is an accurate proxy for different dimensions of cultural distances.

In all the aforementioned literature, it is emphasized that all the measures of cultural and genetic distance are also strongly correlated with geographic distance. Therefore, we must be cautious when interpreting correlations between trade or economic diffusion and cultural or

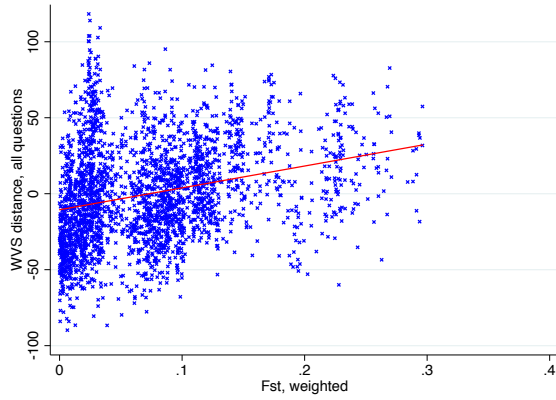
---

<sup>2</sup>They also construct a measure of cultural  $F_{ST}$  which is not publicly available yet

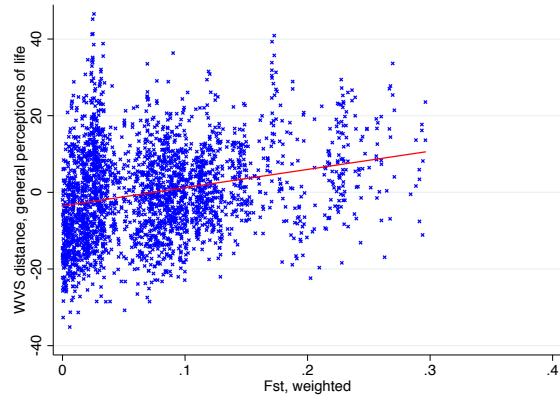
genetic distances. Without controlling for geographic distance, we might as well be capturing standard trade costs (Giuliano et al., 2014). However, geographic distance does not predict all the variation in genetic distances. Figure 3.6 shows the classic multidimensional scaling of the bilateral  $F_{ST}$  matrix. This approach allows the visualization of the relative genetic distance between regions. The colors represent geographic regions of the world: sub-Saharan Africa, North Africa, East Europe, West Europe, South Asia, East Asia, and Latin America. Geographically close areas tend to exhibit lower genetic distance, but there is a large variation. sub-Saharan African regions exhibit much larger genetic distances than East and South Asian ones, for instance.

### 3.3 Conclusion

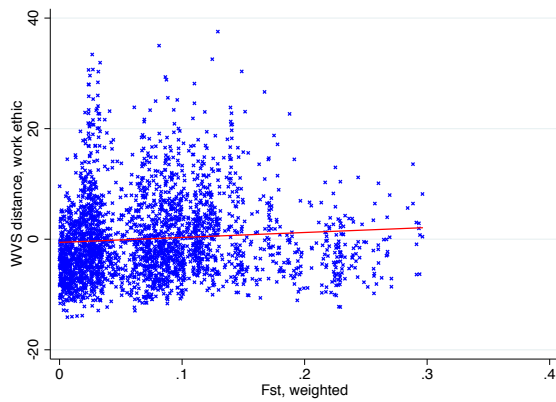
This chapter presented a baseline framework to understand the main measures of population genetics used in Economics: heterozygosity (within-population measures) and genetic distance (between population measures). Although the extrapolation from genetic measures to social conclusions is not always convincing, the analysis of within-population measures has posited interesting and challenging hypothesis that the research in Economics still needs to elucidate. Between population measures of genetic distance are more convincing measures of cultural divergences, if used carefully. There is substantial evidence from Anthropology, Economics, and Population Genetics supporting the evidence that genetic clusters proxy cultural proxies. Chapter 4 of this thesis uses this evidence to study the role of ancestry distance as a predictor of individual economic success in the United States.



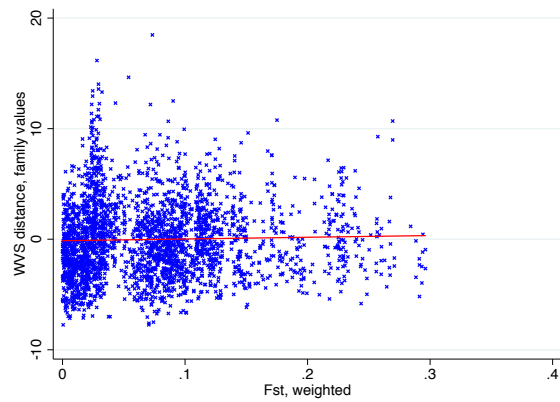
(a) Average of all questions



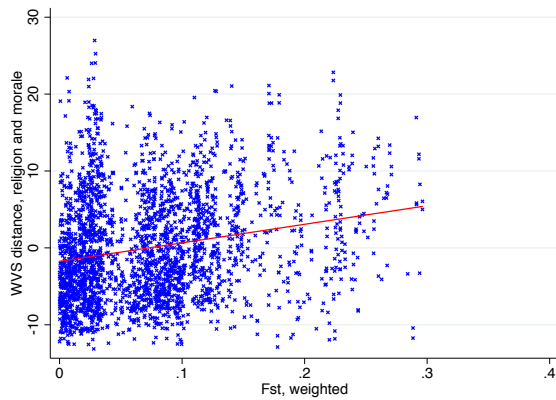
(b) General perceptions of life



(c) Work ethic



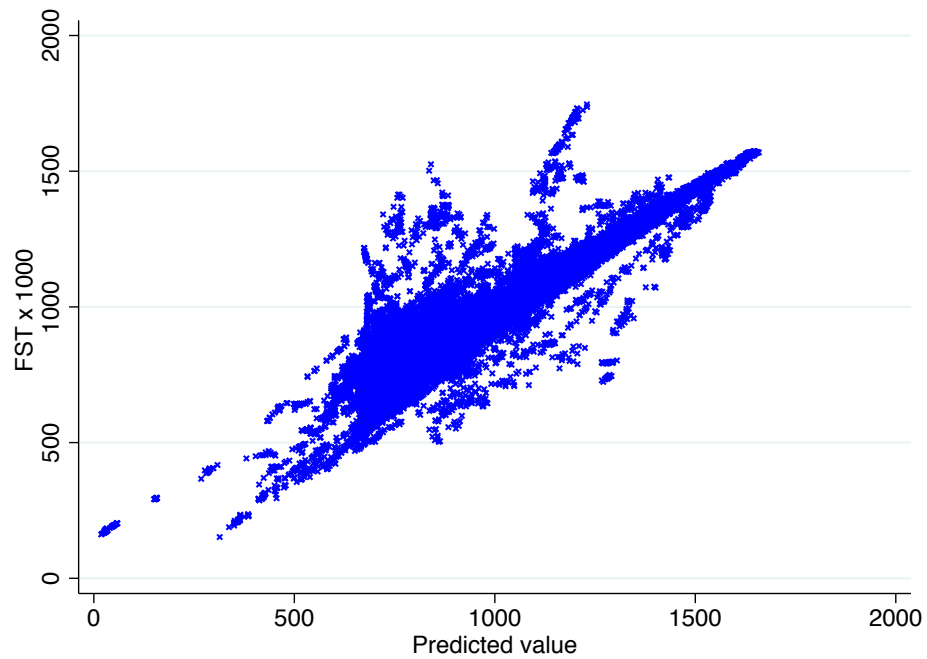
(d) Family values



(e) Religion and morale

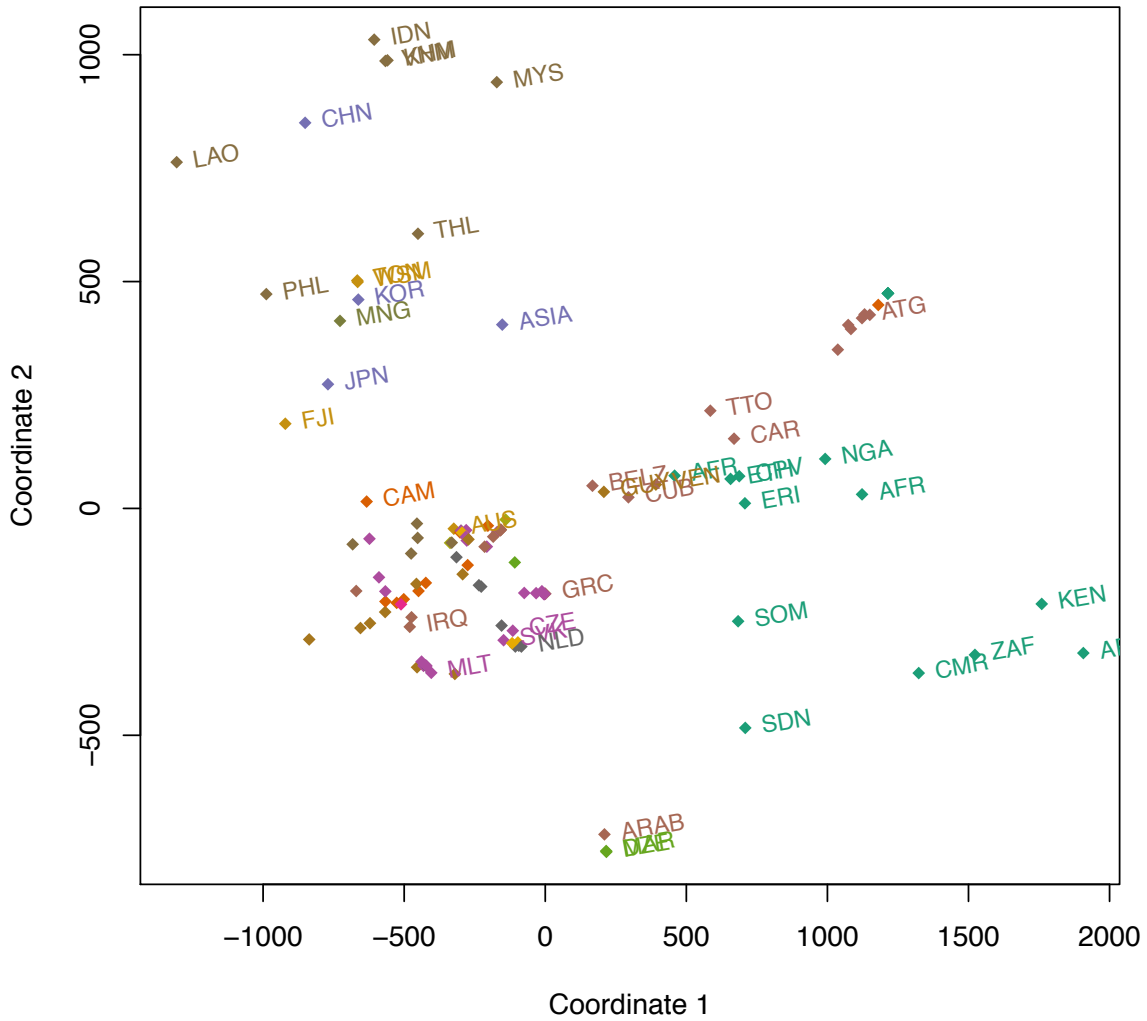
**Figure 3.4** – Correlation between genetic distance ( $F_{ST}$ ) and cultural distance (from the WVS)

**Figure 3.5** – Correlation between  $F_{ST}$  and predicted  $F_{ST}$



**Notes:** This figure shows the correlation between the real  $F_{ST}$  values and the predicted ones in a regression of the  $F_{ST}$  on the different measures of cultural distance from the WVS. The  $R^2$  of this regression is higher than 0.98.

**Figure 3.6** – Multidimensional scaling of the  $F_{ST}$



**Notes:** This figure shows the multidimensional scaling of the genetic distance  $F_{ST}$ . This approach allows the visualization of the relative genetic distance between regions. The colors represent geographic regions of the world: sub-Saharan Africa, North Africa, East Europe, West Europe, South Asia, East Asia, and Latin America



# Chapter 4

## Achieving the American Dream: Ancestry distance, ancestry diversity, and economic success

### Introduction

This paper investigates the role of ancestry differences as determinants of individual economic performance at the local level.

Putnam (2007) states that there is a short-term negative correlation between modern societies' increased diversity and social capital. The research in political economy has produced and extensive literature explaining how increased aggregate cultural or ethnic diversity can hinder collective action because it decreases the provision of public goods (Alesina et al., 1999; Alesina and La Ferrara, 2005; Miguel and Gugerty, 2005), increases the probability of conflict (Collier and Hoeffler, 2004; Montalvo and Reynal-Querol, 2005; Esteban et al., 2012), or decreases the likelihood of social interactions and participation in the community (Alesina and La Ferrara, 2000; Algan et al., 2016). At the cross-country level, cultural differences can decrease the likelihood of technological transfers (Spolaore and Wacziarg, 2009).

Instead of focusing on aggregate levels of diversity, this article investigates the role of bilateral ancestry distance at the individual level as a determinant of economic success. If we can interpret our measure of ancestry distance as a proxy for cultural distance, our



approach sets the micro-foundations explaining the channels through which cultural diversity can impact the economy at aggregate levels. Conditional on the level of diversity, the ancestry difference of a single individual could either be a barrier or a springboard for economic success. On the one hand, cultural similarities ease economic transactions in a daily basis because they restrict the set of possible preferences. In that sense, cultural distance would act as a transaction cost to economic exchanges. On the other hand, cultural differences can be beneficial because different individuals can find new innovation opportunities, or have skills that are complementary and rare in the economy. For instance, Ager and Brukner (2013) show that the massive migration to the USA in the late nineteenth and early twentieth centuries increased the variety of skills, which had in turn a positive effect on growth.

Ashraf and Galor (2013) also suggest that diversity has antagonistic effects income. They claim that at the macro –cross-country– level, there is an inverse U-shaped relationship between genetic diversity and economic output at the cross-country level. They interpret genetic diversity as a measure of within-population diversity, determined thousands of years ago. They attribute the inverse U-shape relation to diversity increasing both innovation and the probability of conflict, meaning that it pulls income in opposite directions. Our results agree with this idea of possible opposite effects of diversity on income, but question the possibility of such long-term persistence at the micro-level.

To obtain individual-level measures of cultural differences, we exploit two sources of variation in the American population. First, the American population has diverse ancestry origins, which cultural traits have persisted over time (Algan and Cahuc, 2010; Ager and Brukner, 2013; Fulford et al., 2015). Second, the recent research by Spolaore and Wacziarg (2009), Desmet et al. (2015), and Spolaore and Wacziarg (2015) suggests that there is an extensive margin in cultural differences. The cultural differences between two pairs of countries depend on the extent of exchanges they have shared in the past. Desmet et al. (2015) also show that this extensive margin in cultural differences can be accurately proxied using population-level genetic distances. For each individual in the American census, we know her location and ancestry. Using the data on pairwise genetic and cultural distances from Spolaore and Wacziarg (2009), and Desmet et al. (2015), as well as information on the ancestry-composition of each area, we obtain a measure of cultural distance for each individual that varies at the location-

ancestry level. Therefore, we can proxy for each individual's differences to the population around her, conditional on her ancestry and her location's aggregate diversity. We then investigate how each person's cultural distance to her local environment affects her economic performance, depending on the local characteristics of her area of residence and her ancestry-inherited social capital.

Economic integration is embedded in the society, and as such, it is the result of an interactive process. If cultural distance can become an economic advantage at the micro level, it will only be so in an environment that is open to such differences. This article explores different societal and individual factors that transform cultural distance as into a relevant predictive factor of economic success. In other words, we explore the interaction effect of cultural distance and societal and individual factors in economic success. These factors are ancestry-inherited levels of trust, ethnic fractionalization, and xenophobia, at both the individual and average-geographic levels.

Our analysis focuses on the heterogeneous effect of cultural distance on income. In particular, regional levels of ancestry-inherited attitudes such as trust or xenophobia, affect the correlation between cultural distance and economic success. Cultural distance affects negatively income living with people coming from ancestries where people are more likely to be xenophobic. On the contrary, we show that cultural distance is positively associated with income in areas where individuals come from trusting ancestries. Finally, we show that newly arrived first-generation migrants tend to earn less, the more culturally distant they are but this effects tends to vanish and converge after a few years spent in the territory.

These results show that at the individual level, culture can indeed act as either a barrier or a catalyst of economic success. However, we challenge the hypothesis of a relationship so deeply rooted that it can persist over thousands of years. We show that individual cultural distance affects economic performance, but this relationship varies depending on a number of factors. In particular, the effect is highly dependent on the local environment's level of social capital.

Our approach overcomes different identification challenges. First, identifying the costs of cultural distance is challenging because cultural differences are usually defined at the country level. Therefore, cultural proximity can also be capturing institutional similarities leading

to trade and economic success. In this article, we bypass this limitation by focusing on the cultural diversity at the *individual level*, in the United States. As noted by Fulford et al. (2015), the United States are a unique laboratory to study the diversity brought to the country by generations of migrants. The American population nowadays is a rich mixture resulting from large migration waves in the recent past.

Second, even within the same institutional setting, identifying the causal effect of cultural diversity on economic outcomes bears the challenge of reverse causality. Economically prosperous areas are more likely to attract migrants from all origins making these locations more likely to be diverse. Since our measure of cultural distance vary at the location-ancestry level, we can control for location-specific effects. The threat of endogeneity at the local level is also present when analyzing the interactive effect of cultural distance with other cultural traits capturing attitudes towards alterity. Other traits could indeed cause changes in trust or openness to others, as suggested by Ashraf and Galor (2013). Our measures of these cultural traits use ancestry-inherited characteristics. Since cultural traits are highly persistent over time (Algan and Cahuc, 2010; Nunn and Wantchekon, 2011; Grosjean, 2014), we use the cultural traits of the respondents' ancestries as a measure of cultural traits that is exogenous to cultural distance.

Third, there are possibly ancestry-specific characteristics that influence economic success nowadays. At the individual level, cultural traits and social capital are transmitted across generations and durably affect the individual's economic performance. This is the case for trust (Guiso et al., 2008; Algan and Cahuc, 2010; Nunn and Wantchekon, 2011), the culture of honor (Grosjean, 2014), or even antisemitism (Voigtländer and Voth, 2012). Since our measure of cultural distance varies at the location-ancestry level, we can also control for ancestry-specific effects.

There is monumental work on the definition of culture in the Social Sciences. According to Lévi-Strauss, culture “ensures the group's existence as a group, and consequently (...) replaces chance by organization”. Culture sets rules organizing the group by imposing a given order upon the contingency of nature (Lévi-Strauss, 1969, p.32). At the individual level, Douglas (1986, p.55) compares cultural norms to “cognitive devices” shared by all the members of a society. These “cognitive devices” are sufficiently rooted in our beliefs, to be

considered as natural or to be unconscious. They translate into categories that organize our views of the world. Our preferences and choices are the result of rational decisions rooted in these “cognitive devices”.

As an example, Douglas (1986, p. 64) uses the division of labor between men and women. In the Western world, the traditional gender differentiation regards women as weaker or, as Douglas puts it, “pretty little things incapable of thought”. In other cultures, women are thought of being only persons capable of doing the hardest tasks. The traditional belief of the Bemuda people in Cameroon is that women are the only people capable of “growing things”, hence the only capable of doing agricultural work. This example clearly illustrates the extent to which culture shapes the organization and the productivity in an economy.

The literature in economics has recently investigated the persistent effect of culture on the economy. Guiso et al. (2006) define culture as *“those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation”*. Therefore, it is in the nature of culture to persist over time. At the individual level, cultural traits and social capital are transmitted across generations and durably affect the individual’s economic performance. This is the case for trust (Algan and Cahuc, 2010; Guiso et al., 2008), the culture of honor (Grosjean, 2014), or even antisemitism (Voigtländer and Voth, 2012). From a theoretical perspective, Bisin and Verdier (2001) model the transmission of preferences as an endogenous process involving the socialization of children in the family, where parents have preferences for transmitting their own culture. In their framework, heterogeneous cultural groups can persist inside a given society. Instead of focusing on the transmission of a single cultural trait, this article explores the role of cultural distance as a whole on economic success.

The literature on cultural distance has mostly explored its role on growth and international trade at the cross-country level. Cultural distance has been shown to increase the relative outcome gap to the technological frontier (Spolaore and Wacziarg, 2009). There is also an extensive international trade literature on the role of cultural distance as an explanatory factor in gravity equations (see van Bergeijk and Brakman (2010), and Lucas (2014, chap. 4) for a detailed review of the literature). We complement this literature because we exploit the Americans’ cultural diversity to understand the role of cultural distance on

economic performance at the individual level.

This paper also contributes to the growing literature on the measurement of cultural diversity. Desmet et al. (2015) and Spolaore and Wacziarg (2015) construct novel indicators of bilateral cultural distances using average answers to the questions of the World Value Surveys. They show that genetic distance is a good summary measure of cultural distances that account for different traits. We complement this literature by considering cultural distance, measured with genetic distances, and cultural diversity as inherited traits and comparing their effects on individual's economic performance.

Another strand of the literature focuses on the role of the separation of society into different groups (class, ethnicities, religion, etc.) as an explanatory factor for conflict or growth. From a theoretical viewpoint, Esteban and Ray (1999) show that the division of a society into different groups has complex effects on conflict. Under certain conditions, the distribution of population into two groups will maximize conflict. From an empirical perspective, there is opposing evidence regarding the effect of culture and identities on conflict and economic performance. On the one hand, Collier and Hoeffler (2004) show that ethnic or religious divisions have little effects on civil conflict compared to economic variables. On the other hand, Montalvo and Reynal-Querol (2005) claim that ethnic polarization, as opposed to ethnic fractionalization, has a robust effect on conflict. Esteban et al. (2012) explore further this debate and they conclude that the effect of polarization (resp. fractionalization) on conflict is heightened by relative publicness (resp. privateness) of the prize. Within countries, aggregate cultural or ethnic diversity can hinder collective action because it decreases the provision of public goods (Alesina et al., 1999; Alesina and La Ferrara, 2005; Miguel and Gugerty, 2005), or decreases the likelihood of social interactions and participation in the community (Alesina and La Ferrara, 2000; Algan et al., 2016).

Our measure of cultural distance allows moving beyond aggregate effects of diversity and question whether cultural differences can be a barrier to the equality of opportunities in the USA. The American Dream describes a society where “*even those who are poor and have limited skills can succeed (...) (the American Dream is) a belief that there is a fair chance of succeeding and ample opportunities to do so. Everyone has a chance, the opportunities are there, and hard work will be rewarded.* (Clark, 2003, Chap.1, p.4). Our results question

this view, as we show that cultural differences can impact economic success, in particular in regions where other individuals come from less trusting and tolerant ancestries.

Major political and historical moments in American history have questioned the ideology of the American Dream. Martin Luther King Jr's historical speech *I Have a Dream* can also be read as condemning the exclusion of African-Americans from the promises of the American Dream, as he wrote:

Let us not wallow in the valley of despair, I say to you today, my friends.  
And so even though we face the difficulties of today and tomorrow, I still have a dream.  
It is a dream deeply rooted in the American Dream  
I have a dream that one day this nation will rise up and live out the true meaning of its creed: "We hold these truths to be self-evident, that all men are created equal"

This article also takes roots in this questioning of the opportunity to achieve economic success across different groups in the population. Recent economic studies have also shown that ancestry has persistent effects on economic outcomes at aggregate levels. The thorough study of Fulford et al. (2015) on the role of ancestry composition on GDP at the county-level in the USA uses census data through History, to show that ancestry-inherited culture, institutions, and human capital are significant explanatory factors of local development in the USA. Similarly, Ager and Brukner (2013) show that in the age of mass migration to the USA, ancestry fractionalization increased income at the county level whereas ancestry polarization decreased it. They explain the different effects because fractionalization captures the effect on income of a more varied pool of skills. On the contrary, polarization would increase political unrest and distortionary taxation. Our results complement this strand of research in two ways. First, our bilateral approach allows to control for ancestry-specific effects. Therefore, we are investigating the role of cultural differences on economic outcomes rather than the effect of certain ancestry-specific persistent cultural traits. Second, instead of finding one immutable relation, we show that at the micro level, the effect of cultural differences are highly dependent on other individuals' levels of trust and xenophobia.

This paper is organized as follows. Section 1 describes the data used in the paper. Section 2 explains the identification strategy, and presents the baseline results. Sections 3

discusses the channels at work. Section 4 presents the robustness checks and additional results. Section 5 concludes.

## 4.1 Data

### 4.1.1 American Community Survey

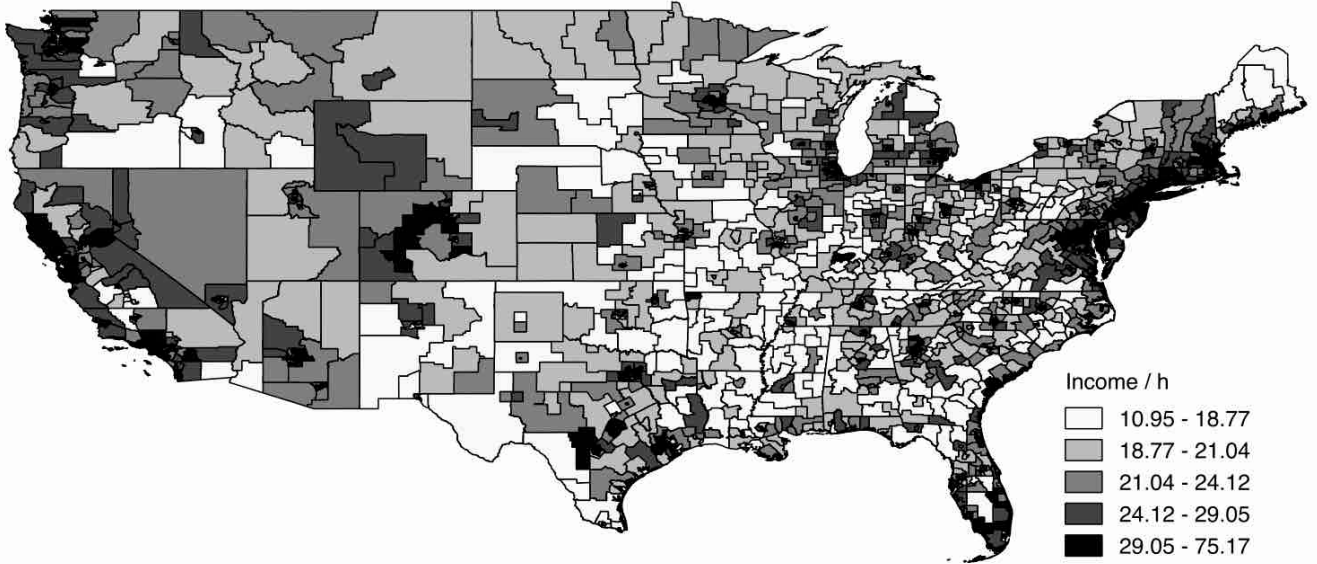
We use the 2010 3-year Public Use Microdata Sample of the American Community Survey (PUMS-ACS) collected by the US Census Bureau. Each record in the data represents a single person and individuals are organized into households. The 2010 3-year PUMS-ACS contains information about 1% of the United-States population for years 2008 to 2010, which is around 3 million observations. For computational purposes, the results come from a representative random sampling of 15% of each year's population. The final sample contains about 1 million observations. The results also passed a check for robustness to the use of another random samples.

For each individual, the ACS reports standard individual characteristics (age, sex, place of birth, etc.), situation in the labor market (income, employment status, etc.), and family situation. Each record also gives information about self-declared ancestry. The list of different ancestries reported is given in the Appendix, in section B.1.1. The survey reports geographic information for PUMA areas. These are geographic units that partition each American state into non-overlapping areas originally defined as containing no fewer than 100,000 people each. Figure 4.1 shows the average hourly income reported in the ACS for in the PUMA areas in the 2008 sample.

### 4.1.2 Genetic Distance as a Proxy for Cultural Distance

#### Definition

Our main proxy for cultural distance, is genetic distance as used by (Spolaore and Wacziarg, 2009). Their measure uses the bilateral genetic distance computed by Cavalli-Sforza et al. (1994) for 42 populations in the world, using 120 gene locus, for which allele frequencies are obtained by population. The gene locus are chosen to represent neutral genes, which means



**Figure 4.1** – Average Hourly Income in PUMA region

genes that did not spread through natural selection but through random drift. Using the mapping of the world’s populations by Alesina et al. (2003), Spolaore and Wacziarg (2009) extend this data into a worldwide bilateral data set. This data is described in more details in the Appendix, section B.1.2.

What is genetic distance exactly measuring? And how can it proxy for cultural distance? The genetic distance used is the fixation index  $F_{ST}$ . For two groups, this indicator shows how much of the variance in genetic markers is due to variance *between* the groups rather than variance *within*.

The  $F_{ST}$  varies between 0 and 1. An index close to 0 will indicate that the two populations interbreed freely, in other words, they can be considered as the same population (most of the variance is due to variance within the populations). The index equals 1 when individuals from the two populations have never interbred, and all the variance is due to differences between the groups. The computation of the  $F_{ST}$  is described in more detailed in the Appendix, section B.1.2.

The longer in time have two populations separated, the more will this separation explain the variation observed. It follows that the fixation index captures how old is the two populations’ most recent common ancestor (MRCA) of the two populations. The closer the index is to 1, the older the MRCA will be.



It is very important to notice that the  $F_{ST}$  does not capture any physical characteristics of the population. The measure captures instead the relationships between two populations using the inherent variability of the genome. Because genes are transferred from generation to generation and are subject to random variations, the differences between two populations in their overall genetic variation (variation in the alleles present in the population for different markers) provide information on the extent of exchanges between the groups. Section B.1.2 in the Appendix shows more precisely how genetic data provides information on population exchanges and their timing.

### Individual-Local $F_{ST}$

The PUMS-ACS asks individuals about their ancestry. For each PUMA unit  $p$ , denote  $s_a^p$  the share of ancestry  $a \in \{1, \dots, A\}$  in the area. Let  $F_{ST}^{a_1, a_2}$  be the genetic distance between ancestries  $a_1$  and  $a_2$ . For each individual living in area  $p$ , from ancestry  $a$ , we construct  $F_{ST}^{a,p}$ , the local average genetic distance to her PUMA area. Equation 4.1 also describes  $F_{ST}^{a,p}$ :

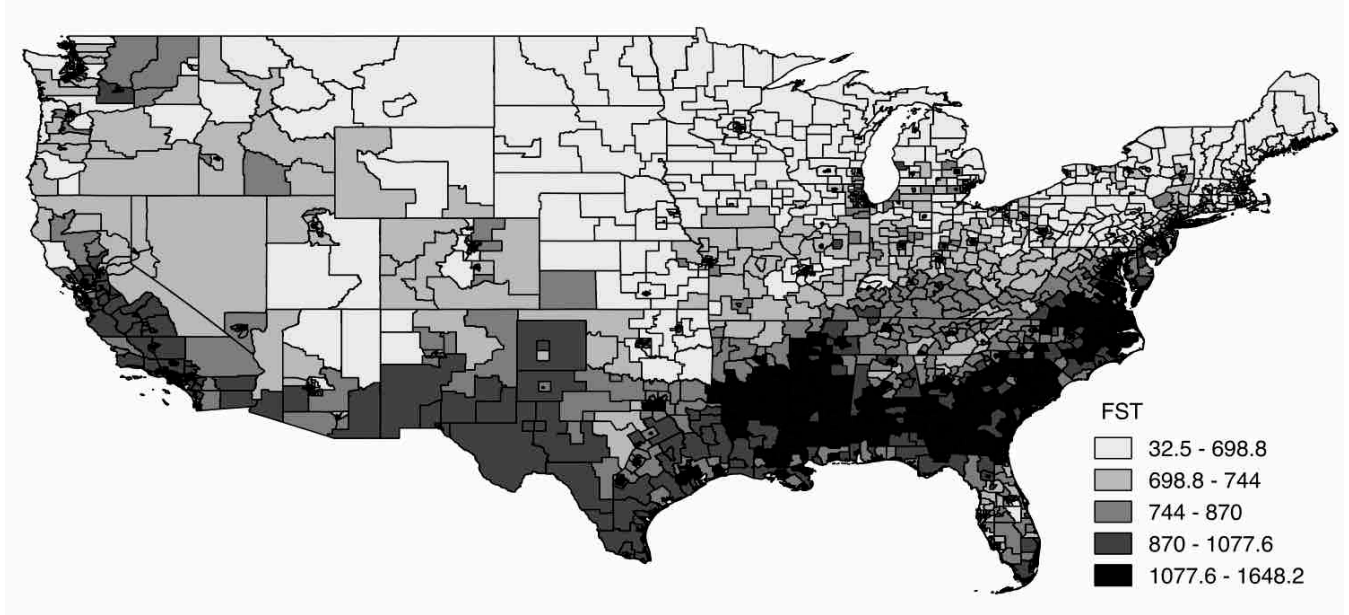
$$F_{ST}^{a,p} = \sum_{l=1}^A s_l^p F_{ST}^{l,a}. \quad (4.1)$$

For simplicity, we refer to the measure  $F_{ST}^{a,p}$  as the  $F_{ST}$  in the remainder of the paper.

Figure 4.2 maps the average  $F_{ST}^{a,p}$  for each PUMA area. The PUMA area with the maximum average is located in Illinois, in the southern of Chicago. The average  $F_{ST}$  there is 0.165. The minimum one is found in a PUMA area located in New Mexico, where the value is 0.003. In comparison, in Spolaore and Wacziarg (2009)'s data, the maximum weighted  $F_{ST}$  between countries is equal to 0.354 and is the one between Papua New Guinea and Zaire. The minimum one is close to 0 and is the one between Sierra Leone and Ghana.

### 4.1.3 Genetic $F_{ST}$ and Cultural Distance

Because the index captures the extent of interbreeding between two populations, it will also capture the extent of cultural exchanges between them. Moreover, it has the advantage of summarizing them into one single measure. Using the World Value Survey (WVS), Spolaore and Wacziarg (2015) construct a series of cultural distance and cultural diversity measures



**Figure 4.2** – Average  $F_{ST}$  in PUMA region

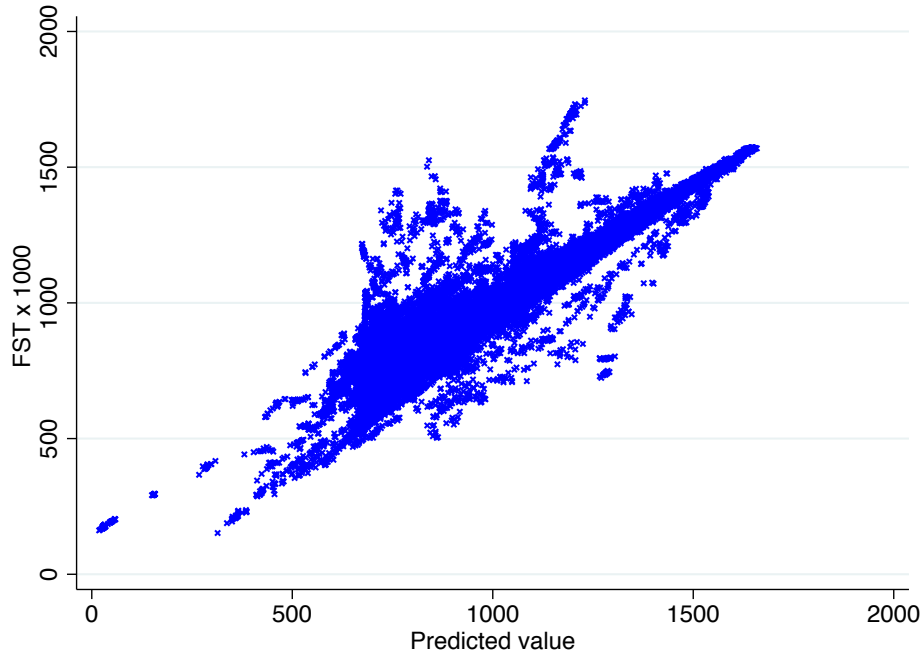
using different traits: language, religion, and values. The WVS’s questions are divided into broader categories that relate to the perceptions of life (like trust, happiness, etc.), work, family values, politics and society, and religion and morale. For each question  $i$ , they construct a measure of cultural distance for a pair of countries  $\{c_1, c_2\}$  that is given by equation 4.2

$$CD = \|s_i^{c_1} - s_i^{c_2}\| \quad (4.2)$$

Where  $\|\cdot\|$  is the Euclidian Norm and  $s_i^c$  is the vector giving the shares of people choosing each of the question’s possible answers.<sup>1</sup> To create summary indices, they standardize the question-specific distances to have zero mean and a standard deviation equal to 1. Finally, they sum the question-specific indices for all the categories in the WVS.

Their analysis of the genetic  $F_{ST}$ ’s predictive power for cultural differences concludes that genetic distance tends to be “*broadly and significantly correlated with a vast range of differences in cultural traits. Thus, while specific correlations with individual sets of traits are typically moderate in magnitude, there is an overall relation between ancestry and cul-*

<sup>1</sup>They also construct a measure of cultural  $F_{ST}$  which is not publicly available yet



**Figure 4.3** – Correlation between the  $F_{ST}$  and its predicted value from other measures of cultural distance

*ture, consistent with a conceptual framework in which a broad range of cultural traits are transmitted with variation across generations over time”* (Spolaore and Wacziarg, 2015).

We also merge these alternative measures of cultural distance to the ACS data. At each location, and for each ancestry, we aggregate these measures in the same way as was done for the  $F_{ST}$ . As an additional robustness test, we regress our individual-local  $F_{ST}$  on these measures of cultural distance. The results are available in the Appendix, Table B.1. The resulting F statistic is large (540,020) and the  $R^2$  is of 0.98. Figure 4.3 shows the correlation between the  $F_{ST}$  and the predicted value  $\hat{F}_{ST}$ , and the scatter-plot is reassuringly close to the 45-degree line. These results confirm Spolaore and Wacziarg (2015)’s results on the validity of the  $F_{ST}$  as a proxy for cultural distance.

In the robustness section of the article, we use these WVS-based measures of cultural distance as alternative explanatory variables. Our results are robust to this change in measure.

### 4.1.4 Diversity

#### Fractionalization

Fractionalization measures how much a society is divided into different groups. More precisely, it captures the probability that two individuals randomly drawn from the population belong to different groups. The index is increasing in the number of groups. Suppose that a society is divided into  $J$  different groups (ethnic, religious, linguistic, etc.). Each group  $j \in \{1, \dots, J\}$  has share  $s_j$ . The index of fractionalization  $FRAC$  is therefore described in equation 4.3.

$$FRAC = 1 - \sum_{j=1}^J s_j^2 \quad (4.3)$$

In this article, we use two different measures of fractionalization: inherited fractionalization and local fractionalization.

Local fractionalization is the ancestry fractionalization at the PUMA level, where the groups are defined by the declared ancestry. This measure captures the contemporary diversity in ancestry at the local level.

By inherited fractionalization, we mean the level of ethnic fractionalization from each individual's country of origin. The data on inherited fractionalization comes from Alesina et al. (2003). It compiles the fractionalization index for ethnic groups across 190 countries respectively. Ethnic fractionalization is computed by identifying different ethnic groups listed in several sources. The sources used are the Britannica (2000), the CIA (CIA) (2000) and the *Ethnic Groups Worldwide* handbook (Levinson, 1998).

The literature has documented a strong correlation between conflict and ethnic fractionalization (Alesina and La Ferrara, 2005; Esteban et al., 2012). Inherited fractionalization is then a measure of the inherited culture of declared ethnic differences and ethnic conflict. Because the culture of violence is highly persistent (Grosjean, 2014), inherited ethnic fractionalization captures cultural attitudes towards alterity. The aim of this indicator is to test whether cultural distance reduces economic opportunities in regions with a strong inherited background of violence against alterity. In other words, we want to look at the interaction between this proxy for inherited ethnic division and cultural distance.

## World Value Survey: Trust and Xenophobia

We compute ancestry-inherited levels of trust and xenophobia from the WVS. The WVS consists of nationally representative surveys conducted in almost 100 countries using a common questionnaire. We use rounds 4 to 6 of the WVS rounds to measure ancestry-inherited cultural traits. Any measure will be the country average of the most recent round for which the country was surveyed. The measure of trust is the country-average of individuals who agree with the statement “most people can be trusted”. This is the same variable used by Algan and Cahuc (2010). The measure of xenophobia is the share of individuals who chose “people who are foreign” when asked to mention “any category that they would not like to have as neighbors”.

### 4.1.5 Summary Statistics

Table 4.2 summarizes the characteristics of individuals in our sample. Column (1) gives the summary statistics for migrants and Column (2) for non-migrants. Column (3) gives the mean difference and its statistical significance. Migrants seem to be different in most observable characteristics, they tend to be younger, less educated, and they are more likely to be unemployed. Because of these differences, section 4.3.1 analyzes separately the differences in the effect of cultural distance between first and further generation migrants.

## 4.2 Cultural Distance and Economic Success

### 4.2.1 Specification

Let  $i$  index individuals,  $a$  index ancestries, and  $p$  index PUMA areas. Equation 4.4 describes our preferred identification equation.

$$y_{iap} = \beta_1 F_{ST}^{ap} + \mathbf{X}_i \boldsymbol{\gamma} + \lambda + \zeta_p + \varepsilon_{ipa} \quad (4.4)$$

In the main specification,  $F_{ST}^{ap}$  is the natural logarithm of the individual-local  $F_{ST}$ .  $\mathbf{X}_i$  is a matrix of individual controls including age, age square, sex, the level of education, marriage status, a binary variable equal to one if the individuals has children, the time spent in the

**Table 4.1** – Summary Statistics ACS

	(1)	(2)	(3)
	Migrant	Born USA	Diff
Female	0.503 (0.002)	0.510 (0.001)	-0.007***
Age	41.444 (0.162)	43.347 (0.106)	-1.902***
Education	16.379 (0.152)	18.577 (0.100)	-2.197***
Time in the USA	19.257 (1.188)	43.347 (0.776)	-24.089***
American Citizenship	0.467 (0.030)	1.000 (0.020)	-0.533***
Children in the house under 6	0.050 (0.001)	0.041 (0.000)	0.009***
Children in the house from 6 to 17	0.128 (0.002)	0.109 (0.001)	0.019***
Is Married	0.646 (0.011)	0.572 (0.007)	0.074***
In Debt 1	0.001 (0.000)	0.001 (0.000)	-0.000***
Unemployed	0.063 (0.002)	0.059 (0.001)	0.004
Yearly Income/1000	33.795 (1.356)	43.171 (0.885)	-9.376***
Observations	244245	1051982	

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports average individual characteristics for individuals in the sample, depending on whether they are first-generation migrants (Column (1)) or not (Column (2)). Column (3) reports the results of a t-test determining whether the difference between Columns (1) and (2) are statistically significantly different from 0.

**Table 4.2** – Summary Statistics of the Diversity Measures

	(1)	(2)	(3)
	Migrant	Born USA	Diff
Fst	951.957 (20.329)	889.657 (13.286)	62.300
<b>Individual-Level Ancestry Characteristics</b>			
Inherited Ethnic Frac	0.396 (0.010)	0.305 (0.007)	0.091***
Inherited Trust	0.262 (0.006)	0.330 (0.003)	-0.067***
Inherited Xenophobia	0.164 (0.002)	0.179 (0.002)	-0.016***
<b>PUMA-Level Ancestry Characteristics</b>			
Regional Frac	0.386 (0.014)	0.338 (0.009)	0.048***
Trust	0.282 (0.007)	0.315 (0.004)	-0.033***
Regional Xenophobia	0.170 (0.003)	0.181 (0.002)	-0.011***
Local Ancestry Frac	0.850 (0.011)	0.848 (0.007)	0.002
Local Ancestry Pol	85.053 (5.491)	78.169 (3.588)	6.884
Observations	186822	902222	

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports average individual and location-level characteristics for individuals in the sample, depending on whether they are first-generation migrants (Column (1)) or not (Column (2)). Column (3) reports the results of a t-test determining whether the difference between Columns (1) and (2) are statistically significantly different from 0.

USA for migrants, its square value, employment status, and declared race (Asian, Black, Hispanic, or White).  $\lambda$  is the year-fixed effect<sup>2</sup> and  $\zeta_p$  is the location fixed-effect. In the most common specification,  $y_{iap}$  is each individual's log hourly income.

### 4.2.2 Baseline Results

Table 4.3 presents the OLS estimates from regression 4.4. Column (1) presents the estimated elasticity of income with respect to the individual's  $F_{ST}$  when no controls are included in the regression. This baseline correlation is negative, which is in line with the previous cross-country level correlations of the effects of cultural distance on international trade.

However, the results from Column (2) show that controlling for individual and puma-level characteristics, as well as including location fixed effects, changes the sign of the effect. With state fixed effects and controls, the correlation between the  $F_{ST}$  and individual income is now positive. This baseline result shows that, at the extensive margin, there is no inherent negative effect between ancestries' differences and individual economic prosperity.

The results from Table 4.3 suggest that, on average, moving from the 25<sup>th</sup> percentile in  $F_{ST}$  to the median will be associated with an increase in income of 183 USD per year (from Column (2)). This result is mostly driven by second and further generation migrants, for whom moving the 25<sup>th</sup> percentile in  $F_{ST}$  to the median will on average be associated with an increase in income of 221 USD per year (from Column (6)). The results are non-significant for first generation migrants (Column (8)). Section 4.3.1 analyzes in more depth the differences in the effect between migrants and non-migrants.

How is it possible that the  $F_{ST}$  would have a strong explanatory power at the macro level to explain relative distances in income to the technology frontier (Spolaore and Wacziarg, 2009), but the effect would be reversed at the micro-level? One possible explanation is that cultural differences have a different effect when considered in an institutionally unified setup. At the micro level, we are studying economic transactions between culturally distant parties that are exempt of institutional risk. Therefore, any frictions in the achievement of economic success due to cultural distance is more likely to be attributed to individual reluctance to

---

<sup>2</sup>To simplify notations the equations do not contain a time index, but they are time-varying as the data contains 3-years of the ACS survey.



**Table 4.3** – Baseline Effect of Fst on Income

	All				Born USA		Migrants	
	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se	(7) b/se	(8) b/se
Log Fst	-0.261*** (0.026)	0.082*** (0.008)	0.149*** (0.011)	0.072*** (0.012)	0.161*** (0.013)	0.099*** (0.017)	0.045*** (0.015)	0.021 (0.016)
Local Fract				-0.034* (0.018)		-0.027 (0.019)		-0.099*** (0.037)
Observations	906,729	906,729	906,729	906,729	729,861	729,861	176,868	176,868
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ancestry FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069	2,069	2,069	2,068	2,068
R2	0.02	0.24	0.25	0.24	0.26	0.25	0.26	0.22

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports summary statistics. Each line reports mean values and standard errors in parenthesis. The first column restrict the sample to individuals born in the USA. The second column to the one of migrants. The last column reports the result of a t-test on the equality of the means.

interact with culturally distant people. In other words, behaviors, rather than institutional barriers are more likely to explain differences in economic success attributable to cultural distance. Section 4.3 of the paper analyzes the interactions between attitudes such as trust and xenophobia and cultural differences to explain economic success.

### 4.2.3 Endogenous Location of Migrants

To correct for possible selection in location, all the regressions control for characteristics of the PUMA area, namely the average per-capita income and the total population. Finally, to capture the likelihood that most prosperous places will attract a more diverse range of the population, Columns (4),(6), and (8) in Table 4.3 and all the subsequent regressions control for local fractionalization (the probability that two individuals randomly drawn from the PUMA area belong to two different ancestries). Finally, all the regressions contain state fixed effects. The inclusion of these controls does not alter significantly the results discussed in previous section.

There is another source of endogeneity at the ancestry level. Consider a location specific location  $p$ . The variable of interest  $F_{ST}^{a,p}$  is given by equation 4.1. For each location consider the random vector  $\tilde{F}_{ST}^p$  of the values taken by  $F_{ST}^{a,p}$  for each location. Given this definition,

there will be an endogeneity problem if:

$$E[\varepsilon'_{pa} \tilde{F}_{ST}^p | \mathbf{X}_i, p] = \sum_{l=1}^A E[\varepsilon'_{pa} s_l^p \tilde{F}_{ST}^{l,a} | \mathbf{X}, p] \neq 0 \quad (4.5)$$

In each location  $p$ ,  $\tilde{F}_{ST}^{l,a}$  is a random vector that attributes the scalar  $F_{ST}$  between ancestry  $a$  and  $l$  to the observation with probability  $p(A = a | \mathbf{X}, p)$ . Therefore, our identification relies on two ignorability assumptions. The first one is considering that  $A | \mathbf{X}, p$  is exogenous. The second one is considering that  $s_l | \mathbf{X}, p, A$  is also exogenous. This section discusses the possible threats and solutions to the identification.

### Selection on Ancestry

The location choice of newly arrived migrants exhibit strong persistence because they are likely to choose regions with a language or economic network that will enhance their opportunities (Crozet, 2004; Bauer et al., 2005, 2006). Therefore, it is possible that  $A | \mathbf{X}, p$  is correlated to region-specific characteristics.

We tackle this problem with two approaches; first, we account for ancestry-specific fixed effects. Second, we account for possible effects of ancestry-networks.

Column (3) in Table 4.3 adds ancestry-fixed effects and controls for the share of individuals from the same ancestry to the baseline estimation. Similarly, we also add the fractionalization in ancestries at the PUMA level, to account for the fact that more prosperous places are more likely to be diverse. Results are shown in Column (4). The results are not changed with respect to Column (2). All the subsequent regressions contain ancestry-fixed effects and control for the average level of ancestry fractionalization.

The heterogeneities in the relationship between the  $F_{ST}$  and income can be visualized in Figure 4.4. It plots the marginal effect of  $F_{ST}^{a,l}$  on the log of income for each different ancestry. The results are ordered increasingly on the size of the estimate. Consistently with Fulford et al. (2015), we find heterogeneity in the effect of cultural distance for different ancestries. The effects range from  $-0.200$  for Syrian ancestry to  $0.563$  for Turkish ancestry. In economic terms, this means that moving from the 25<sup>th</sup> percentile in  $F_{ST}$  to the median will on average decrease (respectively increase) income of 447 USD (respectively 1231) per year

for individuals of Syrian (respectively Turkish) ancestry.<sup>3</sup> There is no obvious ranking of the ancestry-specific effect of cultural differences in income: there are poor and rich countries at both ends of the distribution. Section 4.3 of the paper analyzes the role of attitudes such as trust and xenophobia to explain this heterogeneity.

### **Selection on Individual Unobservables**

Despite our attempts to control for observable individual, location, and ancestry-specific factors, the estimates may still be biased by unobservable factors with cultural distance and economic success. We follow Altonji et al. (2005); Nunn and Wantchekon (2011); Oster (2013) and assess the likelihood that the estimates are biased by unobservables.

A common heuristic for evaluating the robustness of a result to omitted variable bias concerns is to look at the sensitivity of the treatment effect to the inclusion of observed controls. A key issue is the need to make an assumption about the share of the outcome variance would could be explained by observed and unobserved variables together (Oster, 2013). Nunn and Wantchekon (2011) undertake direct calculations based on the theory of Altonji et al. (2005) (AET). In this paper, we follow Oster (2013)'s implementation strategy of the AET method to compare bias due to unobservable characteristics to the one due to observables. We find that it is unlikely that the entire estimated effect of is driven by unobserved variables. The results and the technical details of the estimation are presented in the Appendix Section B.2.1.

## **4.3 Channels**

### **4.3.1 First and Older Generation Migrants**

#### **Generation versus Adaptation**

The literature in migration has shown that strong linkages in migration can have beneficial effects on their economic integration. Migrant networks improve newly arrived migrants' integration in the labor market, because they help them find better jobs (Munshi, 2003),

---

<sup>3</sup>For these calculations, we use the 25<sup>th</sup> percentile level of hourly income and the average American Working Hours as reported by the OECD.



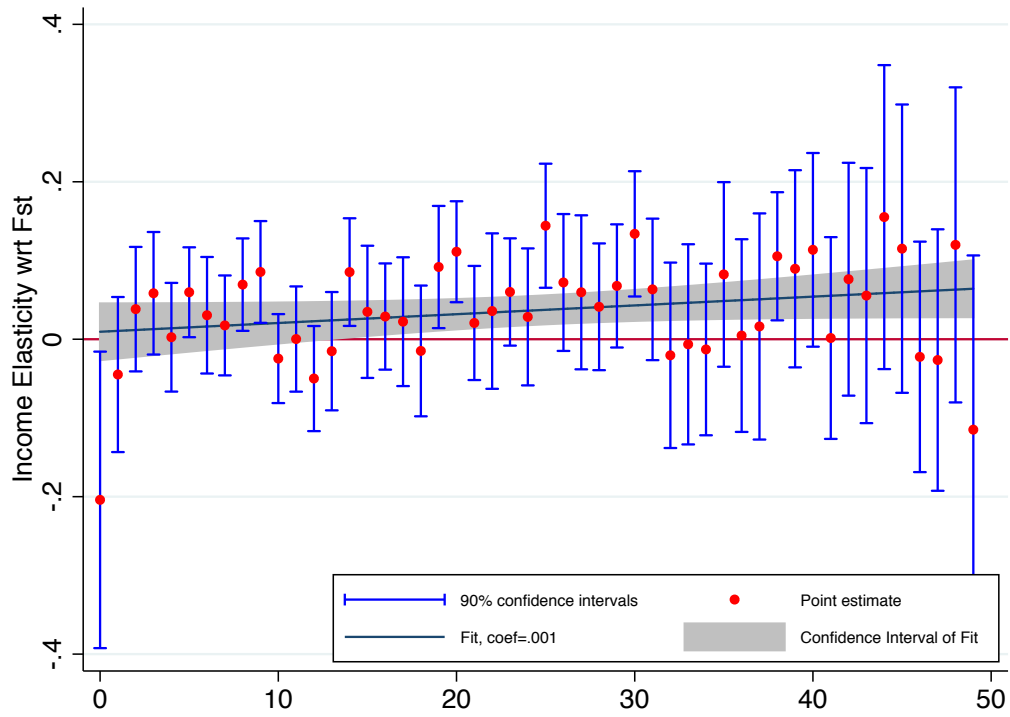
decrease the cost of moving (Carrington et al., 1996), or ease the integration of those who are less proficient in the host country's language (Bauer et al., 2006). In this sense, cultural differences should not have a positive effect on income for first generation migrants.

In this section, we show that the effect of cultural distance on economic performance does indeed depend on whether or not the individual is a first generation migrant. In Table 4.3, we can compare the correlation between the  $F_{ST}$  and income for first generation (Column (8)) and further generation migrants (Column (6)). These columns show the estimates from the baseline when the sample is split between migrants and non-migrants. The baseline correlation is still positive and equal to 0.099, for individuals born in the USA, whereas it is non-significantly different from 0 for foreign-born individuals.

How to interpret the differences in the effects between migrants and non-migrants? A generational hypothesis is that the recent waves of migration might be less educated, or come from less developed countries as compared to the ones of the age of mass migration. This hypothesis would be implicitly stating that more educated ancestries could make better use of their cultural differences. Is it also the case for the predictive power of the  $F_{ST}$  on income? From Figure 4.4, there does not seem to be a clear pattern of more educated or richer ancestries succeeding having the advantage of a positive correlation between  $F_{ST}$  and income. Fulford et al. (2015) also show that ancestry-inherited level of education is not a good predictor of economic performance, and ancestry institutions and GDP are rather weak ones.

Another possible interpretation is that migrants face a sunk cost of adapting to the new local environment. In this case, cultural distance might represent a cost rather than a premium, but it is a sunk cost that individuals can overcome. Figure 4.5 shows the marginal effect of  $F_{ST}$  on economic performance for first generation migrants, depending on how long they have lived in the USA. The marginal effect is indeed negative at the early stages of migrants settlement, but converges towards zero the longer the individuals have been in the territory. This result provides more evidence in favor of the hypothesis according to which cultural differences can indeed be a sunk cost for economic integration, but only in the short-term. Once individuals have adapted, these differences can become an overall positive asset.

**Figure 4.5** – Marginal Effect of  $F_{ST}$  on Income as a Function of Time Spent in the USA, First-Generation Migrants Only



**Notes:** The figure plots the marginal effect of the log of  $F_{ST}$  on the log of income for migrants depending on how long they have been living in the USA. The vertical bars represent the confidence interval of the marginal effects

## The Geographical Distribution of Ancestries

The results from the previous section suggests that cultural distances will only be an advantage for individuals who have already adapted to their local environment. However, this interpretation bears the problem that the PUMA level of aggregation might be too large to gauge the importance of culturally similar networks among migrants.

The previous literature on cultural diversity has suggested that cultural distance might either be a barrier or a springboard for economic success (Ager and Brukner, 2013; Ashraf and Galor, 2013). The argument is that although diversity can hinder cooperation, the diversity of skills induced by cultural differences will benefit development. From the previous sections, it appears that in the particular case of the USA, this positive effect dominates, yielding an average positive correlation between the  $F_{ST}$  and income. For the particular case of first-generation migrants, we expect the effect of cultural distance to be only positive for those who will adapt faster to their new environment. The presence of networks is one factor that is likely to catalyze the integration for the newly-arrived individuals (Carrington et al., 1996).

This section uses a geographical method to measure the extent of cultural networks. Aggregate measures from the ACS at the census-tract level are available from the US Census Bureau. Using this data, we compute measures of ancestry-level spatial segregation to capture the presence of ancestry-level concentrations at the PUMA level. Our chosen measure of spatial segregation is the Entropy index at the PUMA level, which is also the measure used by the Census Bureau.

In a given census tract  $T$ , denote  $n_T$  the total population. For the correspondent PUMA area, denote  $N_T$  and  $N_P$  the total number of census tracts and the total population. Suppose that there are  $A$  possible ancestries. For each census tract  $T$ , in PUMA area  $P$ , we can compute the Entropy score that captures the extent of diversity in the tract. This index is given by  $E_T^P$ :

$$E_T^P = \sum_{a=1}^A -s_T^a \log(s_T^a).$$

Similarly, we can compute the entropy score at the PUMA area:

$$E^P = \sum_{a=1}^A -s_P^a \log(s_P^a).$$

The entropy index is the weighted average deviation of each unit's entropy from the metropolitan-wide entropy, expressed as a fraction of the metropolitan area's total entropy, and it is described in equation 4.6. The entropy index varies between 0, when all tracts have the same composition as the PUMA area, to a high of 1, when each tracts contains one group only (maximum segregation). This index measures how evenly groups are distributed across metropolitan area neighborhoods, regardless of the size of each of the groups.

$$H = \sum_{T=1}^{N_T} \frac{n_T}{N_P} \frac{E^P - E_T^P}{E^P} \quad (4.6)$$

Table 4.4 present the results of the same regressions as shown in Table 4.3, when the  $F_{ST}$  is interacted with a binary variable equal to one if the PUMA area has above-median levels of ancestry segregation. The results are shown for the entire sample (Column (1)), individuals born in the USA (Column (2)) and first-generation migrants (Column (3)).

In regions with high levels of ancestry segregation, cultural distance can turn to be an advantage for first generation migrants, like it is for the rest of the population. Although this is an imperfect measure of the presence of the respondents' networks, the results do provide more supporting evidence for the hypothesis that cultural distance can be an advantage for those who have already adapted to their environment. For first-generation migrants, local networks are likely to fasten this process of integration. In light of this section's findings, all the subsequent generation control for PUMA-level segregation in ancestry, time spent in the USA and a binary variable equal to one if the individual is not a migrant.

If cultural differences only affected economic success through the sunk cost imposed to first generation migrants, we should observe no correlation between the  $F_{ST}$  and economic success among further generation migrants. However, the ancestry-segregation interacts negatively with cultural distance for second and further-generation migrants (Column (2)). This means that everything else equal, individuals will see the positive effect of cultural distance severely diminished when living in highly segregated PUMA areas. One possible explanation is that



**Table 4.4** – Effect of Fst and Local Ancestry Divisions on Income

	(1)	(2)	(3)
	All	Non Migrants	Migrants
	b/se	b/se	b/se
Log Fst	0.108*** (0.023)	0.104*** (0.023)	-0.017 (0.033)
High Ancestry Segregation	0.213* (0.127)	0.268** (0.128)	-0.440** (0.207)
Log Fst x High Segregations	-0.030 (0.019)	-0.038** (0.019)	0.063** (0.031)
Observations	906,729	729,861	176,868
State FE	Yes	Yes	Yes
Ancestry FE	Yes	Yes	Yes
Clusters	2,069	2,069	2,068
R2	0.25	0.26	0.26

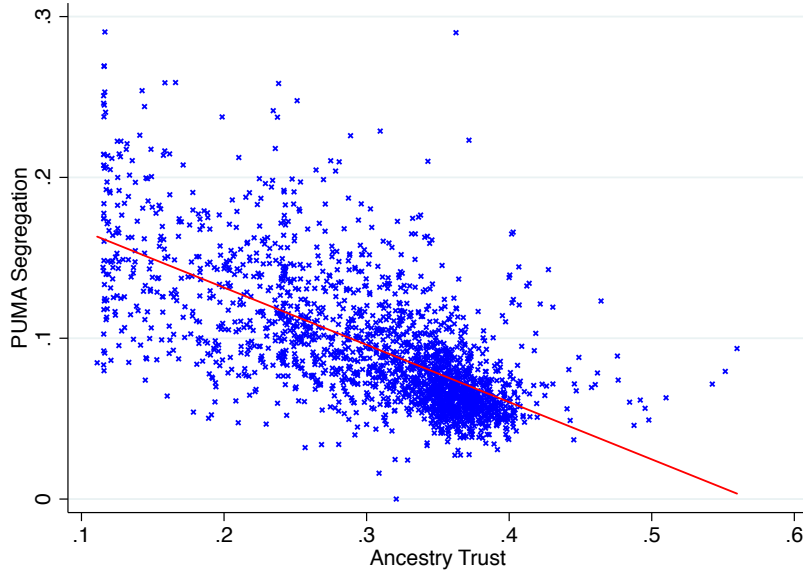
**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual’s earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

the persistence of geographic segregation can also capture other characteristics that in the long run, will counteract the first-hand effect of networks as catalysts of economic success for culturally distant individuals. In particular, individuals from ancestries that are not trusting, or that are closed to alterity, are more likely to segregate spatially. Such characteristics will in turn influence whether culturally distant individuals are likely to succeed. Figure 4.6 shows the negative correlation between ancestry segregation and the average levels of ancestry-inherited trust in the PUMA area. The same results hold for average levels of ancestry-inherited xenophobia and ethnic fractionalization in the PUMA area (Figures 4.7 and 4.8). The reminder of the article aims at determining whether such ancestry-inherited attitudes can explain the economic success of culturally distant individuals.

### 4.3.2 The Role of Attitudes to Explain the Effect of Cultural Distance on Income

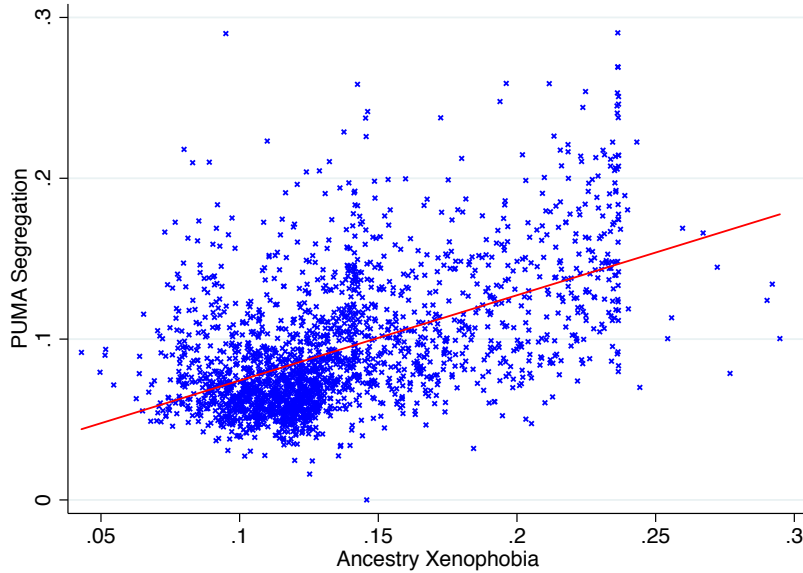
This section aims at determining whether attitudes can trigger the change in the effect of cultural distance on economic performance. We turn to a behavioral interpretation of the effect of cultural distance on economic performance because in an institutionally unified setting, where ancestry and location-specific effects are accounted for, the “cultural barrier”

**Figure 4.6** – Ancestry Segregation and Ancestry-inherited levels of Trust, at the PUMA area



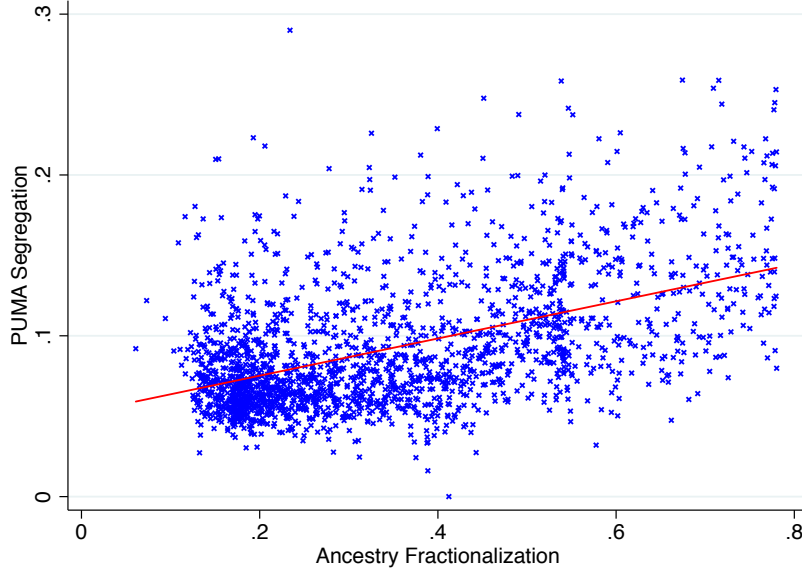
**Notes:** The figure plots the level of ancestry segregation at the PUMA level against the ancestry-inherited trust, averaged at the PUMA level

**Figure 4.7** – Ancestry Segregation and Ancestry-inherited levels of Xenophobia, at the PUMA area



**Notes:** The figure plots the level of ancestry segregation at the PUMA level against the ancestry-inherited xenophobia, averaged at the PUMA level

**Figure 4.8** – Ancestry Segregation and Ancestry-inherited levels of Ethnic Fractionalization, at the PUMA area



**Notes:** The figure plots the level of ancestry segregation at the PUMA level against the ancestry-inherited ethnic fractionalization, averaged at the PUMA level

to economic success is more likely to be attributed to individual reluctance to interact with culturally distant people.

We test for different inherited explanatory factors: fractionalization, trust, and other measures of cultural openness from the World Value Survey (WVS). Each one of these factors is computed as either the “individual” or the “regional” characteristic.

By “individual”, we mean the reported level of trust, fractionalization, or xenophobia of each individual’s ancestry from the WVS. By “regional” characteristics, we mean the average level of the same characteristics of the ancestries residing in the PUMA area of the respondent. We interact each one of these factors the log of  $F_{ST}$ . Sections 4.3.2 and 4.3.2 present the results.

An implicit assumption for using the ancestry characteristics from the WVS is that cultural traits are inherited throughout generations. This is why we refer to them as being “inherited”. Bisin and Verdier (2001), and Algan and Cahuc (2010) support theoretically and empirically (respectively) the validity of such assumption. Cultural traits do persist over time among older generation migrants. The advantage of using this ancestry-inherited

approach is that we capture the variation in attitudes that is exogenous to the economic characteristics and ancestry-composition of the respondents' locations.

### Specification

$$y_{ija} = \beta_1 F_{ST}^{k,a} + \mathbf{High Div}_i \beta_2 + \mathbf{High Div}_j \times F_{ST}^{k,a} \beta_3 + \mathbf{X}_i \boldsymbol{\gamma} + \lambda + \zeta_a + \eta_k + \theta_y + \varepsilon_{ija} \quad (4.7)$$

In the main specification,  $F_{ST}^{k,a}$  is the natural logarithm of the individual-local  $F_{ST}$ . For each of the measures of cultural and genetic diversity described in the previous section, we construct a binary variable that will be equal to one if the ancestry's measured diversity is higher than the median. **High Div** is a vector containing these binary variables.  $\mathbf{High Div}_j \times F_{ST}^{k,a}$  is the vector of interactions between the high diversity binary variables and the individual-local  $F_{ST}$ .

### Ancestry-Inherited Individual Characteristics

The measures of **High Div<sub>j</sub>** used in this section are binary variables equal to one if the individual's ancestry has a level of trust, ethnic fractionalization, or xenophobia higher than the median for the different ancestries. As stated before, in an institutionally unified setting, only behaviors and attitudes towards alterity can explain how cultural differences interact with economic success. The aim of this section is to determine whether individual attitudes can explain the role of cultural differences on economic success.

Income, cultural differences, and attitudes towards alterity are likely to be endogenously determined. Throughout History, xenophobic movements have indeed taken root among the most deprived populations, feeding from the angst of finding a scapegoat for their condition. Ancestry-inherited attitudes are a strong predictor of individual contemporary attitudes that are exogenous to income and the contemporary cultural composition of the region (Algan and Cahuc, 2010).

Table 4.5 reports the results. The  $F_{ST}$  is only a significant –positive – predictor of economic individual success when individuals come from highly trusting ancestries (Column (2)). The effect is negative, albeit not statistically significant, for individuals from highly fractionalized ancestries, and close to zero for those from highly xenophobic ancestries. The results

are not very robust, suggesting that individual attitudes, in particular trust, might explain the role of the  $F_{ST}$  on income, but not in a sufficiently powerful way. It is therefore necessary to seek for alternative sources of explanation.

From Column (2), the predicted elasticity of income with respect to the  $F_{ST}$  for individuals from highly trusting ancestries is 0.081 and the baseline elasticity is 0.054. Moving from the 25<sup>th</sup> to the 50<sup>th</sup> percentile of  $F_{ST}$  brings 60 USD more per year for those from highly trusting ancestries. The effect is therefore also small in economic terms and not robust to the simultaneous inclusion of all the interactions (Column (4)).

Because the results found in this section are small in economic terms and not robust, next section explores the role of the same cultural characteristics, *when considering the average among individuals living in the same area as respondents*. The economic success of individuals is the result of social interactions, and as such it can be hindered by cultural barriers (language, different norms, or prejudice), but also enhanced if individuals can make innovative use of their cultural differences, skills and knowledge diversity (Ager and Brukner, 2013). Therefore, the attitudes of the surrounding individuals should matter at least as much as the respondents' to explain the heterogeneous predictive power of the  $F_{ST}$  on income.

### **Ancestry-Inherited Regional Characteristics**

In this section, the measures of **High Div<sub>j</sub>** are binary variables equal to one if the average measures of diversity in the PUMA area is higher than the median. The measures used are still ancestry-inherited cultural attitudes (ethnic fractionalization, trust, and xenophobia), but we consider the average of those measures in the PUMA area. The aim of this section is to test whether attitudes at the society level determine whether cultural differences affect economic outcomes.

Table 4.6 reports the results. The  $F_{ST}$  is a significant negative predictor of economic individual success when individuals live in areas with people from highly ethnically fractionalized (Column (1)) or xenophobic ancestries (Column (3)). On the contrary, the  $F_{ST}$  is a positive predictor when individuals live in areas with people from highly trusting ancestries (Column (2)).

**Table 4.5** – Effect of Fst and Ancestry-Inherited Individual Characteristics on Income

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Log Fst	0.077*** (0.010)	0.054*** (0.012)	0.073*** (0.012)	0.099*** (0.027)
<b>Fractionalization</b>				
High Ancestry ethnic	0.077 (0.088)			-0.071 (0.130)
Log Fst x High Ancestry ethnic	-0.010 (0.013)			0.012 (0.019)
<b>Trust</b>				
High Trust		-0.199** (0.090)		-0.173 (0.135)
Log Fst x High Trust		0.027** (0.013)		0.023 (0.020)
<b>Xenophobia</b>				
High Xenophobia			-0.040 (0.087)	-0.121 (0.107)
Log Fst x High Xenophobia			0.005 (0.013)	0.018 (0.016)
Observations	866,051	866,051	866,051	866,051
State FE	Yes	Yes	Yes	Yes
Ancestry FE	No	No	No	No
Clusters	1,969	1,969	1,969	1,969
R2	0.24	0.24	0.24	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

The baseline predictive power of the  $F_{ST}$  on income is now positive, but it is not robust to the simultaneous inclusion of all the interactions (Column (4)). The inclusion of all the interactions also greatly diminishes its magnitude. This result suggests that there is no baseline effect of cultural distance on income, and that any effect found is results from its interaction with the society’s attitudes towards alterity.

The predicted elasticity of income with respect to the  $F_{ST}$  for individuals living in areas with people from highly fractionalized, highly xenophobic, and highly trusting ancestries are , respectively, -0.050, 0.110, and -0.081 (Columns(1) to (3)). In economic terms, this means that ,on average, moving from the 25<sup>th</sup> to the 50<sup>th</sup> percentile of  $F_{ST}$ , brings 100 USD less per year for those living in areas with people from highly fractionalized ancestries, 226 USD more for those those living in areas with people from highly trusting ones, and 161 USD less for those for those those living in areas with people from highly xenophobic ancestries.

When all the interactions are included simultaneously, the ancestry-inherited regional levels of trust and xenophobia remain statistically significant categories to understand the heterogeneous effect of the  $F_{ST}$  on economic performance, whereas fractionalization is not. Overall, the results are in line with the hypothesis that the role of cultural differences will depend on the society’s attitudes towards alterity. In the USA, the baseline effect of cultural differences is the one of a springboard for economic success, but only in areas where others are open to such differences.

## 4.4 Robustness

### 4.4.1 Alternative Outcomes

We have shown that cultural distance affects economic success, in particular when interacted with proxies for the society’s degree of openness to alterity. Average level of trust in the society pulls the effect upwards whereas the average degree of xenophobia pushes it downwards. In this section, we explore alternative measures of economic integration, namely the probability of being self-employed and the income of the self employed. The aim of this analysis is to test whether our results are robust to such alternative measures and consistent whether our interpretation of the results is in line with these new outcomes.

**Table 4.6** – Effect of Fst and Ancestry-Inherited Regional Characteristics on Income

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Log Fst	0.096*** (0.022)	0.007 (0.015)	0.112*** (0.021)	0.016 (0.027)
<b>Fractionalization</b>				
High Region ethnic	0.355** (0.162)			-0.258 (0.176)
Log Fst x High Region ethnic	-0.050** (0.024)			0.040 (0.026)
<b>Trust</b>				
High Trust		-0.764*** (0.124)		-0.708*** (0.136)
Log Fst x High Trust		0.109*** (0.018)		0.101*** (0.020)
<b>Xenophobia</b>				
High Xenophobia			0.510*** (0.136)	0.451*** (0.125)
Log Fst x High Xenophobia			-0.081*** (0.020)	-0.072*** (0.019)
Observations	906,729	906,729	906,729	906,729
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.



There is a substantial work in the role of self-employment as a vector of economic integration. On the one hand, self-employment among migrants is often a way to achieve economic success and to integrate in the host country Borjas (1986); Zimmerman and Constant (2004). However, the persistence of self-employment does not necessarily signal better equality of opportunities in the labor market. First, self-employed individuals do not necessarily earn more than employees (Blanchflower, 2004). Second, the recent findings in development economics have shown that self employment among the more deprived communities is often the result of lack of other employment opportunities. Using the words of Banerjee and Duflo (2011), these individuals are “reluctant entrepreneurs”.

Do the patterns of employment observed in the USA confirm the hypothesis that cultural distance affects economic success depending on the society’s attitudes towards alterity? Table 4.7 shows the result of running regression (4.7) when the outcome is the probability of being self-employed (Column (1)), in which case this is a logit regression), the income of the self employed (Column (2)) and the income among the employed (Column (3)). First, the baseline effect of the  $F_{ST}$  is that it increases the probability of being self employed. Second, living with other individuals from trusting ancestries does decrease the probability of being self-employed for culturally distant individuals (Column (1)). Moreover, among self-employed individuals, living with other individuals from trusting ancestries tends to increase income for culturally distant individuals, whereas the opposite holds for living with other individuals from ancestries with higher levels of xenophobia (Column (2)). These results are in line with the hypothesis that the effect of cultural differences on economic integration are highly dependent on the society’s attitudes towards alterity.

As opposed to the previous results, the interaction between the  $F_{ST}$  and the average level of ancestry-inherited ethnic fractionalization decreases the probability of being self employed. However, the results on income among the self employed is not statistically significant. One possible explanation for such inconsistent results for ethnic fractionalization can be due to the fact that it is not exactly a behavioral measure per se and can therefore capture a wide range of behaviors resulting from diverse factors from the culture of violence to being used to cultural diversity.

**Table 4.7** – Effect of Fst and Ancestry-Inherited Regional Characteristics on the Probability of Being Self Employed and the Income of the Self-Employed

	(1) P(Self E.) b/se	(2) Self E. Income b/se	(3) Employed Income b/se
main			
Log Fst	0.263** (0.102)	-0.034 (0.056)	0.026 (0.025)
<b>Fractionalization</b>			
High Region ethnic	1.728** (0.768)	-0.800* (0.452)	0.119 (0.175)
Log Fst x High Region ethnic	-0.260** (0.115)	0.122* (0.068)	-0.020 (0.026)
<b>Trust</b>			
High Trust	2.780*** (0.609)	-0.896** (0.359)	-0.587*** (0.139)
Log Fst x High Trust	-0.408*** (0.091)	0.129** (0.053)	0.085*** (0.021)
<b>Xenophobia</b>			
High Xenophobia	-0.720 (0.548)	0.709** (0.290)	0.579*** (0.122)
Log Fst x High Xenophobia	0.114 (0.082)	-0.108** (0.043)	-0.089*** (0.018)
Observations	1,149,732	106,091	798,058
State FE	Yes	Yes	Yes
Ancestry FE	Yes	Yes	Yes
Clusters	2,064	2,064	2,064
R-sq		0.16	0.28
Pseudo R-sq	0.06		

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports Logit (Column (1)) and OLS (Columns (2) and (3)) estimates. The unit of observation is an individual. The dependent variable is the probability log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

### 4.4.2 Alternative Measures of Cultural Distance

We perform the same regressions as in Table 4.6, using the alternative WVS-based measures of cultural distance computed by Spolaore and Wacziarg (2015) and described in section 4.1.3. All the results are reported in the Appendix. The measures focus on religious distance (Table B.2), distance in the general perceptions of life (Table B.3), work ethic (Table B.4), family values (Table B.5), religion and morale (Table B.6), and an average of all the previous categories (Table B.7).

Overall, the results are consistent with the interpretation of cultural distance as having a changing role on income that will mostly depend on the society's attitudes towards alterity. As opposed to the results from Table 4.6, the baseline effect of cultural distance is more likely to be negative or not statistically different from zero with these measures. However, the interaction terms show that average levels of ancestry-inherited trust (respectively ethnic fractionalization, or xenophobia) consistently push the effect of cultural distance upwards (respectively downwards). The effect of trust is the most robust of all, with all measures except one showing a positive and statistically significant interaction effect between cultural distance and ancestry-inherited levels of trust on economic success.

### 4.4.3 Polarization

Ager and Brukner (2013) show that polarization is historically negatively associated with county-level GDP whereas fractionalization is positively associated with it. They explain the different effects because fractionalization captures the effect on income of a more varied pool of skills. On the contrary, polarization would increase political unrest and distortionary taxation. Our analysis focuses on the interaction effect between cultural distance at the individual level and economic success. One possibility is that cultural distance at the individual level would be capturing the negative effect of polarization at the aggregate level.

Polarization captures the degree of antagonism of a given society. We use the measure of polarization proposed by Esteban et al. (2012) (from Esteban and Ray (1994)) and described in equation 4.8.

$$POL = \sum_{j_1=1}^J \sum_{j_2 \neq j_1}^J s_{j_1}^2 s_{j_2} d_{j_1 j_2} \quad (4.8)$$

We suppose that a society is divided into  $J$  different groups, and each group  $j \in \{1, \dots, J\}$  has share  $s_j$ ,  $d_{j_1 j_2}$  is the distance between the groups. The distance is a measure capturing the heterogeneity of antagonisms between the groups. For instance, it can be measured with cultural distance or the difference in resources between the groups. Unlike fractionalization, polarization declines with the fragmentation of a society into more groups and is globally maximized with a bimodal distribution. We compute the measure of local polarization with the  $F_{ST}$  measure for  $d_{j_1 j_2}$ . The groups are defined by the ancestries reported in the ACS. Using a simplified measure of polarization (considering that  $d_{j_1 j_2}$  is constant for all country pair).

Table B.8 in the Appendix adds  $POL$  as a control in the baseline regression. The results are not changed and the effect of  $POL$  is negligible in magnitude.

## 4.5 Conclusion

What can the diversity-rich North America tell us about the cultural costs of economic success?

This article explores the role of individual cultural distance on income, using the genetic distance as a proxy for cultural distance. We show that cultural distance has heterogeneous predictive power. The results vary depending on different categories. In particular, culturally distant individuals living in regions with other individuals from more trusting ancestries or less xenophobic ones are more likely to be economically successful. First generation migrants seem to be less likely to success the more culturally distant they are, but this effect vanishes as time spent in the USA increases.

In *How Institutions Think*, Douglas (1986, p.55) compares cultural differences to “something like Williamson’s account of transaction costs except that in this case all the advantages lie with joining the corporate effort to make fonder analogies to the work and very little advantage lies with the privateer working under his own flag”. In this article, we stand with the view that cultural differences can be thought of as sunk costs that are not deeply and

immutably rooted in our societies. For first generation migrants, cultural distance can act as an “intangible barrier” to economic success, even in a society as diverse as the United States’. However, this cost vanishes and converges towards zero for those who have stayed longer in the territory.

In some regions, cultural distance is even associated with higher levels of income, even when controlling for ancestry-specific fixed effects. This is more likely to be the case in regions populated with individuals from highly trusting ancestries, and in less ancestry-segregated areas (for second and further-generation migrants). On the contrary, the correlation is more likely to be nil or negative in regions where individuals come from highly ethnically fractionalized or xenophobic ancestries. These results go in line with a non-static view of the role of culture as a barrier or catalyst of economic success. Economic integration is embedded in the society, and as such, it is the result of an interactive process. If cultural distance can become an economic advantage at the micro level, it will only be so in an environment that is open to such differences.

Our results shed a new light to Putnam’s words that “successful immigrant societies create new forms of social solidarity and dampen the negative effects of diversity by constructing new, more encompassing identities. Thus, the central challenge for modern, diversifying societies is to create a new, broader sense of ‘we’” (Putnam, 2007). The economic success of each individual is deeply rooted in these “forms of social solidarities”. Whether culturally different individuals succeed or fail will, at the end, depend on how her environment is ready to accept her talents. In our interpretation, a society fearing that culturally different people will not succeed might very well self-fulfill their prophecy if the fear turns into rejection.

Our research challenges the static view that cultural differences are necessarily an obstacle to economic performance in the long-run. We also challenge the hypothesis of a relationship so deeply rooted that it can persist over thousands of years. According to our analysis, the success of a culturally diverse society is tightly linked to its capacity to accept others. Future research should aim at examining whether such attitudes are subject to change as a result of different public policies and the resulting effect of cultural diversity on income.

# Appendix A

## Appendix to Part 1

### A.1 Data

#### A.1.1 Data on Mission Centers Location and Investments

The *Geography Atlas of Protestant missions*, published in 1903, was the result of an extensive work of localizing all the missionary stations around the world and reporting all the activities they were investing in. Of all the reports, conducted in 1896, 1903, 1911 and 1925, this one (1903) is the most precise and extensive version as investments, denomination, number of students, teachers (both native and foreign) and missionaries are reported for each mission localized on the maps.

The Atlas contains maps of all the regions in the world and locates all the missions active in 1903. Figure A.2 provides an example of these maps. All the missions are uniquely identified in a statistical index providing detailed information on the type of infrastructure available and the number of workers and students. Figure A.1 shows a section of this statistical index.

- A = Anti-opium society.
- a = Anti-foot-binding society.
- B = Bible, or tract and book-room.
- b = Blind school, or speial work for the blind.

**A**

**Abaco**, see **Great Abaco Island**  
**Abadiyeh**, 12-Palestine Inset. — **FFMA** (1899) m w 3-N 2-n 3-x=c s 2-v D  
**Abasa**, ? Gold Coast, Plate 14. — **WMS** m 64-Nn 64-O 684-x=10-c 7-s 8-v  
**Abbotabad**, 11-D1. — **CEZMS** (1894) 3-W n=160 senana pupils  
**CMS** (1899) 2-m w (p) (l) N 9-x=D  
**Abetify**, 14-D7. — **B** (1876) 2-m 2-w 20-N n 17-O 731-x=s 15-v K (Abetifi)  
**Abeth**, very near **Beakleen**, 12-Palestine Inset. — See **Mt. Lebanon**, work of **FM**  
**Abokuta**, 14-ET. — **CMS** (1846) 2-m w 2-W 24-N 5-n 1122-x=c 17-v T D L  
**SBC** (1856) m w 2-N n 2-O 26-x=c s v  
**WMS** 2-m 30-Nn 22-O 264-x=7-c 7-s 3-v  
**Aberdeen**, 13-C1. — **BFSS** (1874) 2-m  
**Abetifi**, see **Abetify**  
**Abo**, see **Aboa**  
**Aboa**, 14-F8 Inset. — **MGB** (1900) m=c v (Abo)  
**Abokobi**, see **Agbogba**  
**Abors**, corner of Tibet, China and Assam, north of **Dibru-garh**, 10-H1. — **MFA** (1900) 2-m (l)=D  
**Aburah**, 14-D7. — **WMS** m 49-Nn 75-O 545-x=10-c 10-s 5-v  
**Aburi**, 14-D7. — **B** (1847) 3-m 2-w W 14-N 4-n 9-O 716-x=10-v K  
**WMS** 3-m W 93-Nn 18-O 644-x=15-c 13-s 14-v h  
**Accra**, see **Accra**  
**Accra**, or **Akra**, 14-D7. — **NBC** (1900) m 50-x=c s  
**WMS** 3-m 34-Nn 30-O 532-x=4-c 7-s 2-v  
**Achampatti**, 12-E6. — **CMS** N  
**Achona**, ? Ind. Ter., 3 and Inset. — **FMH** (1887) N 26-x=c s  
**Acra**, **Akka**, or **Accho**, 12-Palestine Inset, 13-L7. — **CMS** (1890) 2-m 2-w 4-W (p)=D H (Acra)  
**Acton Homes**, 18-E5. — **WMEA** 29-N 12-O 297-x=2-c 3-s 2-v

**Aleppo**, 12-K9, 13-M6. — **PCEJ** (1895) 2-m (p) 3-N n=s B v D  
**Alert Bay**, 2-16. — **CMS** (1878) 2-m 2-w 2-W N n 22-x h I  
**Alessandria**, 13-EF4. — **MEN** N O 61-x=s  
**Alexandria**, 12-L10, 13-Ks, 15-G1. — **AJM** No statistic  
**BFSS** 2-m 3-N=B  
**CSJ** 4-m 3-w 3-W=4-v  
**DAK** (1857) 13-W=H  
**EMB** (1898) 7-m  
**IU** (1898) m w=home of rest  
**LSPCJ** No statistics  
**NAM** (1892) 2-m 2-w 2-W N n=c v  
**UP** (1857) 2-m w 2-W=B  
**WMS** m 2-N 45-x=s  
**Algeciras**, 13-B6, 7. — **SSM** (1890) m w N n=c v  
**Algiers**, 13-D6, 14-E1. — **BFSS** (1882) m w 6-Nn=B  
**FSEJ** m  
**NAM** (1892) 2-m 2-w 2-W (l)=2-s  
**FB** m w W  
**PCFI** No statistics  
**SMF** (1887) m w=c s  
**TM** (1888) 5-W  
**Aligarh**, 11-FG5. — **CMS** (1863) m w 3-W 21-N 10-n 8 82-x=c 10-v  
**MEN** 4-m 2-w W 14-N 16-n 720-x=2-c 32-s S 15-v I  
**Aliwal North**, 17-G7, 18-B7. — **PMMS** (1870) 2-m 2-w 5 9-O 1188-x=c s 2-v 2-T I t  
**Allahabad**, 10-A2. — **BFSS** m=B  
**CMS** (1859) 9-m 6-w 12-N n 300-x=c 5-v T  
**MEN** 2-m w 9-N 9-n 67-x=c 27-s S 13-v f o  
**FN** (1836) 4-m 2-w 5-W (p) 18-N 8-n 2-O 150-x=3-c 6-v 3-h D H  
**WU** (1898) 2-W N 31-n=6-s B 14-v Rescue work, a: special services for beggars  
**YMCA** (1900) m=Y  
**ZEM** 6-W N 19-n=5-v h o 260 senana pupils  
**Alleppe**, see **Alleppey**  
**Allanah**, 12-D2. — **WMS** (1816) — (l) 10-v 2-s 1-c 1-o

**Notes:** This Figure is a reproduction of part of Statistical index from the *Geography Atlas of Protestant missions*. Abbreviations are used in this index for the different characteristics and investments of the missions.

**Figure A.1** – Part of the Statistical index from the *Geography Atlas of Protestant missions*, 1903

- C = College, or university.
- c = Church building, or chapel.
- D = Dispensary.
- d = Deafmute school, or work for deaf and dumb.
- F = Foundling asylum, or nursery mission.
- f = Female helper training class.
- H = Hospital.
- h = High, or boarding school.
- I = Industrial school.
- i = Insane asylum.
- K = Kindergarten.
- L = Leper asylum, or special leper work.

APPENDIX A. APPENDIX TO PART 1

- l = Literary worker.
- M = Medical, or nurses' class, or school.
- m = Male foreign missionary.
- N = Native male teacher, catechist, or pastor.
- n = Native female teacher, or religious helper.
- Nn = Total number of native workers of both sexes.
- O = Outstation.
- o = Orphanage.
- P = Printing press.
- p = Physician.
- R = Refuge for opium victims.
- r = Rendezvous for training new missionaries.
- S = Society of young people.
- s = Sunday-school.
- T = Theological class, or school.
- t = Temperance society.
- V = Village, or day-school.
- W = Woman who is unmarried, or is a widow.
- w = Wife of a missionary.
- X = White Cross Society.
- x = Christians in native church.
- Y = Young Men's Christian Association.



### A.1.2 Afrobarometer Data

We use data from the Afrobarometer survey, rounds 3 and 4 (resp. 2005 and 2008), which are a comparative series of national surveys on attitudes towards democracy, markets, and civil society in Africa. Surveys contain representative information at the individual level of attitudes towards political and social outcomes as well as individual evaluations of living standards. Table A.1 gives a brief description of all the variables from the Afrobarometer used in our analysis.

A precise description of the survey, variables and sampling methods is available on the Afrobarometer website, <http://www.afrobarometer.org>. Similarly, all the questionnaires used for the construction of the dataset are available on the <http://www.afrobarometer.org/survey-and-methods/questionnaires> website. Table A.1 describes the questions used to construct the outcomes in the analysis.

### A.1.3 Controls

#### Historical Characteristics

- **Cities 1400:** indicator variable that equals one if there was a city located on the land inhabited by each ethnic group. **Source:** Nunn and Wantchekon (2011) using data from Chandler (1987) on the location of African cities with more than 20,000 inhabitants in year 1400.
- **Distance to 1400 City:** the distance is computed by the authors using ArcGIS. **Source:** The information on the location of cities in 1400 is from Nunn and Wantchekon (2011).
- **Distance to 1800 City:** the distance is computed by the authors using ArcGIS. **Source:** The information on the location of cities in 1800 is from Nunn and Wantchekon (2011).
- **Distance to Catholic mission in 1889:** the distance is computed by the authors using ArcGIS. **Source:** The information on the location of missions in 1889 is from

Nunn (2009). He geocoded the original map from Béthune (1889).

- **Explorer contact:** indicator variable that equals one if a European explorer traveled through land historically occupied by the ethnic group (the variable captures exploration routes between 1768 and 1894). **Source:** Nunn and Wantchekon (2011) using information on the location of railway lines in the first decade of the twentieth century from Company (1911).
- **Initial population density:** The average population density estimate in the 18<sup>th</sup> Century in a buffer of 10 km around each location. From the HYDE 3.1 database.
- **Malaria ecology of the land:** Malaria stability index. The index takes into account the prevalence and type of mosquitoes endemic to a region, their human biting rate, their daily survival rate, and their incubation period. It has been constructed for 0.5 degree by 0.5 degree grid cells globally (Kiszewski et al., 2004).
- **Railway contact:** indicator variable that equals one if any part of the railway network was built on land historically inhabited by the ethnic group. **Source:** Nunn and Wantchekon (2011) using information on the location of railway lines in the first decade of the twentieth century from Company (1911).
- **Slave exports, per capita:** estimates of the number of slaves taken from each ethnic group disaggregated at the ethnicity level. These estimates were constructed by combining data on the total number of slaves shipped from all ports and regions of Africa with data on the slaves' ethnic identities. The estimates cover Africa's transatlantic and Indian Ocean slave trades between 1400 and 1900. **Source:** Nunn (2008).

### Geographic Characteristics

- **Annual precipitation level:** the annual precipitation level data refers to total yearly precipitation. The model used to build this data has been applied considering the baseline period 1961-1990. A detailed description of calculation procedures used to build this data can be found in the GAEZ v.3.0 Global Agro Ecological Zones - Model Documentation (Fischer et al., 2000). **Source:** Global Agro-Ecological Zones data.

- **Average elevation:** in meters. **Source:** Nunn (2008) using data from Parker (1997).
- **Distance to the capital:** the distance is computed by the authors using ArcGIS.
- **Distance to the coast:** the distance to the coast is identified by the authors using the Proximity Utility in ArcGIS.
- **Number of agriculture growing days per year:** number of days during the year when temperature regime and moisture supply are conducive to crop growth and development. This period is also termed the “length of the growing period” (LGP). It is a proxy for the agroclimatic potential productivity of land. The LGP is determined based on prevailing temperatures and water balance calculations for a reference crop. **Source:** Global Agro-Ecological Zones data.
- **Population density in 2005:** population density of the respondent’s neighborhood. **Source:** Afrobarometer (2005).
- **Share of land within 10 km of water:** in percent. **Source:** Nunn (2010) using data from the Digital Chart of the World (DCW).
- **Suitability for rainfed crops:** rainfed suitability has been calculated for nine crop groups that are important to most farming systems in developing countries, namely cereals, fibre crops, oil crops, pulses, roots and tubers, stimulants, sugar crops, tree fruits and vegetables. The algorithm examines in each gridcell all the crop types belonging to a particular crop group. Among these it determines the crop type that maximizes agronomic suitability. **Source:** Global Agro-Ecological Zones data.

#### A.1.4 Newspaper Market

**Newspaper Data** The cross-country newspaper dataset we build relies on various sources:

- **Tudesq (1995):** *Feuilles d’Afrique: étude de la presse de l’Afrique sub-saharienne.*
- **Daubert (2009):** *La presse écrite d’Afrique francophone en question: essai nourri par l’essor de la presse française.*

- The **Standing Conference on Library Material on Africa** (SCOLMA): it provides an inventory of the available African newspapers in archives collected in 1973.
- The *Directory of African Media* (Maja-Pearce, 1996): digitized by the authors.
- The *African book world and press: a directory* (Zell, 1980): digitized by the authors.
- The *Willings press guide* (Redman and Group, 1993, 2003, 2012): digitized by the authors (one issue every ten years).
- **ICON**: using a web-spider, we collect all the data from the *Icon database* which provides information (first and last date published, title and frequency) on newspapers around the world. This information comes from all the records available in a precise set of institutions: the British Library, the Center for Research Libraries, Harvard University, the Library of Congress, the Library and Archives of Canada, the New York Public Library, the New York State Library, the Online Computer Library Center Inc., the University of Florida, the University of Illinois at Urbana-Champaign, the University of North Carolina at Chapel Hill, the University of Southern California, the University of Washington.
- **Readex**: We collect the summary data from the *Readex database*. Over several decades, Readex has published many of the most widely used collections of primary source research materials in academic libraries, first as Readex Microprint Corporation and since 1984 as a division of NewsBank. These digital collections include information for 40 African newspapers from 1800 to 1922.<sup>1</sup>

We merge the information from all these datasets/books together to build a consistent database of African newspapers through history. This dataset is a panel from 1800 to 2012 with the number of newspapers published each year in each sub-Saharan country and their date of creation. One caveat of this dataset is that some newspapers (especially small local

---

<sup>1</sup><http://www.readex.com>

newspapers) may be missing. However we think that it provides an interesting overview of the state of the newspaper market and of its evolution over time.

**Suggestive evidence** Figure A.3 shows the cross-country correlation between the number of newspapers that have existed and the average distance of the cities surveyed in the Afrobarometer to the closest historical mission settlement with a printing press. The figure displays a negative correlation between the average distance to the printing press and the total number of newspapers recorded. Figure A.4 shows the correlation between the date of publication of the first newspaper and the average distance of the cities surveyed in the Afrobarometer to the closest historical Protestant mission station with a printing press. We observe a positive correlation between the distance to the printing press and the publication date of the first newspaper: the closer a location is to a historically attested printing press, the sooner the first newspaper is recorded.

APPENDIX A. APPENDIX TO PART 1



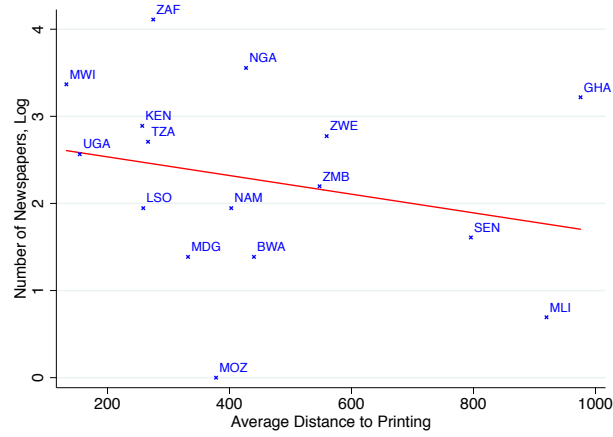
Figure A.2 – Plate Representing all the Missionary Stations in 1903, West Africa (*Geography Atlas of Protestant missions*, 1903)

**Table A.1** – Variables from Afrobarometer, Description

Variable	Original Question in the Survey	Construction
Read News	How often do you get news from the newspapers? Every day, a few times a week, a few times a month, less than once a month, never, don't know	BV, true if at least once a month or more
Listen News	How often do you get news from the radio? Every day, a few times a week, a few times a month, less than once a month, never, don't know	BV, true if at least once a month or more
Watch News	How often do you get news from the TV? Every day, a few times a week, a few times a month, less than once a month, never, don't know	BV, true if at least once a month or more
Trust	How much do you trust the other [Ghanaians/Kenyans/etc.] Not at all, just a little, somewhat, a lot, don't know	BV, true if somewhat or more
Register	Understanding that some [Ghanaians/Kenyans/etc.] were not able to register as voters for the [20xx] elections, which statement is true for you? You were registered to vote, You did not want to register, You could not find a place to register, You were prevented from registering, You were too young to register, Did not register for some other reason, Don't Know/Can't remember.	BV, true if "You were registered to vote"
Turnout	With regard to the most recent nationale election, which statement is true? You were not registered or too young, you voted in the election, you decied not to vote, you could not find the polling station, you were prevented from voting, you did not have time to vote, you did not vote for some other reason, you could not find your name in the register, don't know	BV, trie if "you voted in the elections"
Actions	Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: attend a community meeting; go together with others to raise an issue; attend a demonstration or protest march.	BV, true if says yes to any of the options
Education	Level of education from none (0) to graduate studies (9)	DI
Health	[Is there a health clinic] in the primary sampling area or within easy walking distance?	BV, true if yes

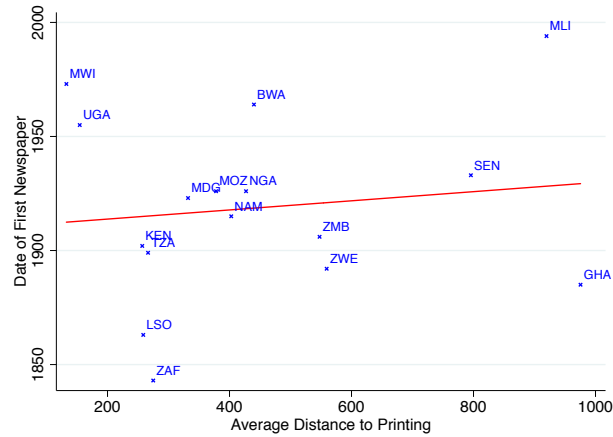
**Notes:** DI refers to *Discrete Index*, BV to *Binary Variable*

APPENDIX A. APPENDIX TO PART 1



**Notes:** The x-axis represents the country average distance of Afrobarometer to the closest mission with printing press. The y-axis represents the country total number of newspapers in the dataset.

**Figure A.3** – Distance to the Printing Press and Number of Newspapers



**Notes:** The x-axis represents the country average distance of the villages in the Afrobarometer to the closest mission with a printing press. The y-axis represents the publication date of the first newspaper in the country.

**Figure A.4** – Distance to the Printing Press and Publication Date of the First Newspaper

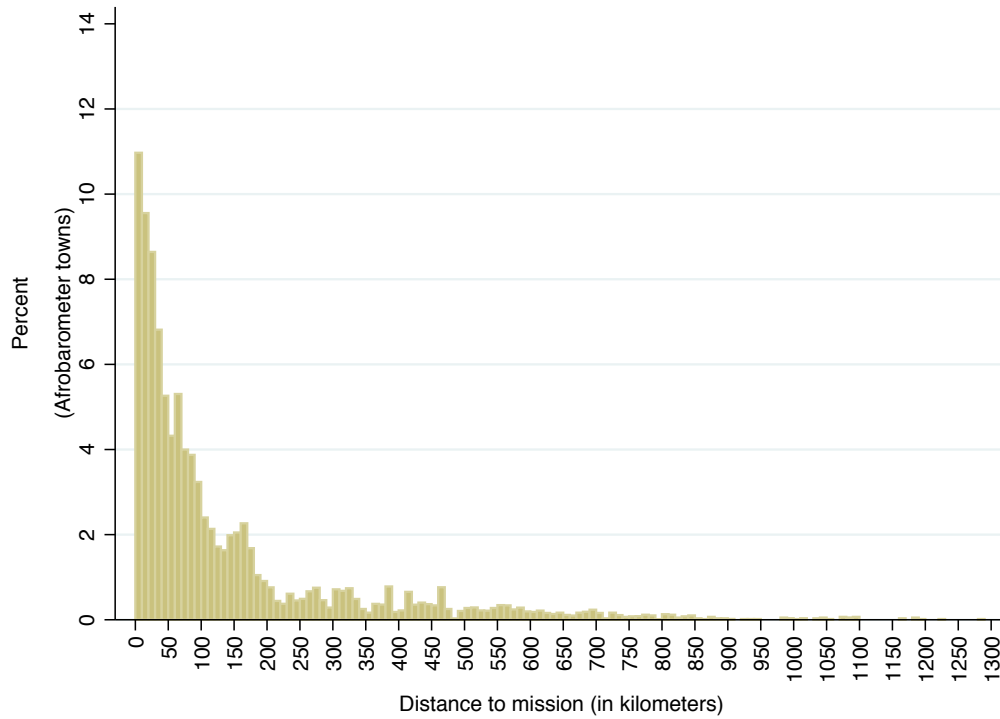


### A.1.5 Descriptive Statistics

**Table A.3** – Summary statistics: Distance of the Afrobarometer towns to the closest mission and to the closest mission with a printing press.

	Mean	sd	Median	Min	Max
Distance to mission	140	184	68	0.16	1,286
Distance to printing press	439	308	362	1.27	1,676
Observations	47,456				

**Notes:** The Table gives summary statistics for the distance of Afrobarometer towns to Protestant missions. Values for distances are in kilometers.



**Notes:** The Figure plots the distribution of the distance of the Afrobarometer towns to the closest mission (with 10 km bins). Values for distances are in kilometers.

**Figure A.5** – Distribution of the distance of the Afrobarometer towns to the closest mission

**Table A.4** – Summary statistics: Afrobarometer data

	All sample	100km
<b>Outcomes: Social and civic capital</b>		
News	0.33 (0.47)	0.38 (0.49)
Trust	0.42 (0.49)	0.38 (0.49)
Turnout	0.72 (0.45)	0.72 (0.45)
Registration	0.80 (0.40)	0.81 (0.39)
Actions	0.24 (0.43)	0.23 (0.42)
<b>Other outcomes</b>		
Education	3.11 (2.02)	3.41 (1.92)
Clinic	0.57 (0.50)	0.56 (0.50)
<b>Controls</b>		
Female	0.50 (0.50)	0.50 (0.50)
Age	36.35 (14.54)	36.49 (14.71)
Observations	47,456	29,411

**Notes:** The Table gives summary statistics for the characteristics of the individuals surveyed in the Afrobarometer. Column 1 presents the statistics for the entire sample. Column 2 presents the statistics for the individuals living in towns located less than 100 km away to the closest mission.

**Table A.5** – Historical and Geographical Characteristics of Printing Presses With and Without a Publication Record

	(1) No Printing	(2) Printing	(3) Diff
<b>Geographical Characteristics</b>			
Number of Growing Days (%)	51.393	66.541	-15.148 (9.434)
Suitability for Rainfed Crops	5.750	4.000	1.750** (0.704)
Average Precipitation per Day (over year)	3.607	3.928	-0.320 (0.838)
Distance to 2000 city (100km)	2.497	2.773	-0.275 (0.855)
Distance to the Coast (100km)	1.307	3.828	-2.521** (1.102)
Malaria Ecology	7.618	12.302	-4.685 (3.609)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.058	0.032	0.026 (0.047)
Railway Contact	0.083	0.125	-0.042 (0.122)
Explorer Contact	0.250	0.312	-0.062 (0.179)
Initial Population Density / 100	17.611	10.872	6.739 (8.418)
Distance to 1400 City (100km)	6.097	7.624	-1.527 (1.678)
Distance to 1800 City (100km)	8.963	12.010	-3.047 (3.203)
Observations	12	16	28

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table compares geographical and historical characteristics of missions with a printing press depending on whether they had a publication record in 1923 or not. Column 1 presents the results for missions with no publication record. Column 2 presents the results for missions with a publication record. In Column 3 we perform a t-test on the equality of means (standard errors in parenthesis). Variables are described in the Appendix.

**Table A.6** – Investments of Printing Presses With and Without a Publication Record

	No Printing	Printing	Diff
Arrival Date	1866	1876	-10 (10)
Bible Society	0.583	0.188	0.396 (0.229)
Number of Native Workers	2.583	1.188	1.396 (2.615)
Total Population	431.167	400.188	30.979 (229.650)
Schools	2.333	1.125	1.208 (0.937)
Number of Students	408.333	389.000	19.333 (227.537)
Schools per Student (%)	2.522	0.471	2.051 (1.439)
Teachers per Student (%)	12.435	24.002	-11.566 (15.442)
Health Facilities	1.667	1.312	0.354 (0.679)
Physicians per Capita (%)	0.045	1.633	-1.588 (1.808)
Health Facilities per Capita (%)	0.496	3.657	-3.160 (3.598)
Observations	12	16	28

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table compares investments' of missions with a printing press depending on whether they had a publication record in 1923 or not. Column 1 presents the results for missions with no publication record. Column 2 presents the results for missions with a publicatin record. In Column 3 we perform a t-test on the equality of means (standard errors in parenthesis). Variables are described in the Appendix.

## A.2 Matching: Additional tables

## A.3 Using selection on observables to assess the bias from unobservables: Technical details and results

Suppose there is a set of unobservable explanatory variables  $W'$ . Since these variables are unobserved, they are not included in regression (1.1). The *proportional selection assumption* (PSA) states that  $\delta C_{WX} = \frac{C_{W'X}}{V_{W'}}$ .  $X$  is the treatment variable (here distance to the printing press),  $W$  is the set of observed covariates,  $C_{W'X}$  is the covariance of  $W'$  and  $X$  and  $V_{W'}$  is the variance of  $W'$ .  $\delta$  is a measure of the relationship between  $C_{WX}$  and  $C_{W'X}$ . The PSA assumption is key in the approach as it states that the relationship between the observed covariates  $W$  and the treatment  $X$  is informative about the relationship between the unobserved covariates  $W'$  and  $X$ , from which the bias is coming.

Consider the following three regressions:

$$\text{News} = \gamma X + W + W' + \varepsilon_{\max} \quad (\text{M-max})$$

$$\text{News} = \xi X + M + \varepsilon_1 \quad (\text{M-1})$$

$$\text{News} = \Lambda X + W + \varepsilon_2 \quad (\text{M-2})$$

Let  $R_{\max}$  be the R-squared of the full model regression (M-max).  $R_2$  is the R-squared of the regression (M-2) including all the observed covariates.  $R_1$  is the R-squared of regression (M-1) including only a restricted set of covariates  $M$ .  $M$  is a set of observed controls that do not have a related unobserved component and are orthogonal to  $W$  and  $W'$  (Oster, 2013).

According to Oster (2013), under the PSA and when  $\delta$  is close to one,  $B(\delta) = \delta \frac{(\xi - \Lambda)(R_{\max} - R_2)}{R_2 - R_1}$  is (i) equal to the unobserved bias if  $\delta = 1$ ; (ii) a close upper bound on the bias if  $\delta < 1$ ; (iii) and a close lower bound on the bias if  $\delta > 1$ .<sup>2</sup>

We can compute the bias due to unobserved variables from the movements in the treatment effect due to the added controls using the ratio  $B(\delta)$ . However, to compute the ratio

---

<sup>2</sup>If  $W$  is selected randomly from  $\{W, W'\}$ , then  $\delta = 1$ . If  $W$  is the most important set of controls from  $\{W, W'\}$  then  $\delta < 1$ .

**Table A.7** – Estimation of the propensity score (logit)

	(1)	(2)	(3)
	PP	School	Health
	b/se	b/se	b/se
<b>Geographic Characteristics</b>			
Number of Growing Days (%)	-0.028 (0.031)	-0.012 (0.011)	0.021 (0.018)
Suitability for Rainfed Crops	-0.237 (0.190)	0.057 (0.109)	-0.009 (0.093)
Average Precipitation per Day (over year)	2.284** (0.961)	-0.007 (0.078)	-0.404 (0.535)
Distance to 2000 city (100km)	-0.202 (0.252)	-0.038 (0.106)	0.283*** (0.104)
Distance to the Coast (100km)	0.023 (0.106)	-0.317*** (0.095)	-0.129** (0.055)
Malaria Ecology	0.123** (0.062)	0.022 (0.028)	-0.024 (0.030)
<b>Historical Characteristics</b>			
Slave Exports, per capita	-3.137 (12.498)	-0.549 (0.387)	0.485** (0.211)
Railway Contact	-0.971 (0.709)	-0.384 (0.602)	-1.310*** (0.457)
Explorer Contact	-1.204 (1.270)	1.357* (0.703)	0.936* (0.517)
Initial Population Density / 100	-0.045** (0.017)	-0.011*** (0.004)	-0.003 (0.005)
Distance to 1400 City (100km)	0.150 (0.100)	-0.059 (0.078)	0.009 (0.042)
Distance to 1800 City (100km)	-0.048 (0.052)	-0.041 (0.052)	-0.056*** (0.019)
<b>Mission Characteristics</b>			
Arrival Date	-0.005 (0.021)	0.001 (0.013)	0.012 (0.009)
Bible Society	-2.045 (1.504)	1.177 (1.304)	0.373 (0.852)
Native Workers	-0.372** (0.184)	-0.070 (0.083)	-0.385*** (0.105)
Total Mission Population	0.308*** (0.102)	0.082 (0.069)	0.376*** (0.106)
Number of Students	-0.308*** (0.102)	-0.081 (0.069)	-0.376*** (0.106)
Physician	2.610 (1.621)	-0.276 (0.877)	3.860** (1.570)
Observations	311	393	441
Country FE	Yes	Yes	Yes
Clusters (Society)	41	45	53
Pseudo R2	0.507	0.383	0.505

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports the first stage of the matching strategy. This first stage estimates the propensity score at the mission-level. The propensity score is the estimated probability of observing the mission given the level of mission of the individual, including whether a

APPENDIX A. APPENDIX TO PART 1

**Table A.8** – Propensity score matching: Balancing test (Printing Press)

	No Printing Press	Printing Press	Diff/se
<b>Geographic Characteristics</b>			
Number of Growing Days (%)	66.652	60.232	6.420 (7.118)
Suitability for Rainfed Crops	4.400	4.654	-0.254 (0.540)
Average Precipitation per Day (over year)	5.007	3.812	1.196 (1.581)
Distance to 2000 city (100km)	2.788	2.684	0.105 (0.540)
Distance to the Coast (100km)	2.072	2.655	-0.583 (0.940)
Malaria Ecology	11.945	10.212	1.733 (2.832)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.479	0.046	0.433 (0.309)
Railway Contact	0.200	0.115	0.085 (0.103)
Explorer Contact	0.240	0.231	0.009 (0.121)
Initial Population Density / 100	13.683	14.746	-1.063 (6.303)
Distance to 1400 City (100km)	7.898	7.234	0.664 (1.334)
Distance to 1800 City (100km)	10.872	10.966	-0.094 (2.433)
<b>Mission Characteristics</b>			
Arrival Date	1872.160	1871.760	0.400 (6.439)
Bible Society	0.200	0.308	-0.108 (0.158)
Native Workers	0.000	1.923	-1.923 (1.401)
Total Mission Population	434.520	443.077	-8.557 (224.025)
Number of Students	424.920	427.000	-2.080 (222.933)
Physician	0.200	0.308	-0.108 (0.136)
Observations	51		

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table compares geographical and historical characteristics of missions with a printing press and the missions matched using the propensity score. Column 1 presents the results for missions without a printing press. Column 2 presents the results for missions with a printing press. In Column 3 we perform a t-test on the equality of means (robust standard errors in parenthesis). Variables are described in more details in the text.



**Table A.9** – Propensity score matching: Balancing test (School)

	No School	School	Diff/se
<b>Geographic Characteristics</b>			
Number of Growing Days (%)	58.684	55.155	3.529 (4.348)
Suitability for Rainfed Crops	5.212	5.000	0.212 (0.298)
Average Precipitation per Day (over year)	3.606	4.271	-0.665 (1.220)
Distance to 2000 city (100km)	2.407	2.269	0.138 (0.293)
Distance to the Coast (100km)	2.275	1.533	0.742* (0.373)
Malaria Ecology	10.162	7.456	2.705 (1.584)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.376	0.025	0.350* (0.152)
Railway Contact	0.242	0.185	0.057 (0.068)
Explorer Contact	0.136	0.210	-0.074 (0.063)
Initial Population Density / 100	9.634	17.899	-8.266 (8.218)
Distance to 1400 City (100km)	6.899	7.868	-0.969 (0.815)
Distance to 1800 City (100km)	11.228	12.228	-1.000 (1.405)
<b>Mission Characteristics</b>			
Arrival Date	1874.561	1873.463	1.098 (3.690)
Bible Society	0.197	0.235	-0.038 (0.079)
Native Workers	0.773	3.481	-2.709 (1.809)
Total Mission Population	510.030	663.358	-153.328 (178.616)
Number of Students	503.273	650.988	-147.715 (177.331)
Physician	0.076	0.123	-0.048 (0.054)
Observations	147		

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table compares geographical and historical characteristics of missions with a school and the missions matched using the propensity score. Column 1 presents the results for missions without a school. Column 2 presents the results for missions with a school. In Column 3 we perform a t-test on the equality of means (robust standard errors in parenthesis). Variables are described in more details in the text.

APPENDIX A. APPENDIX TO PART 1

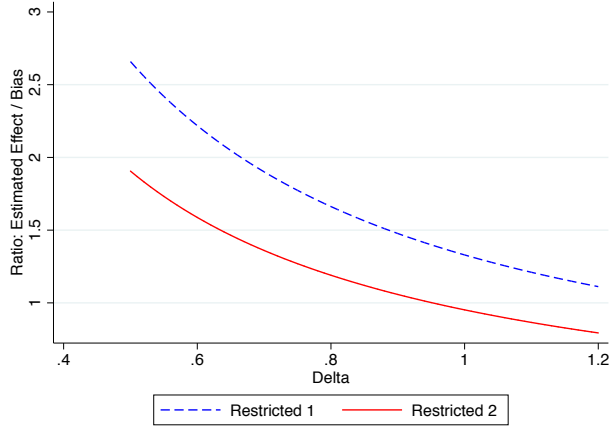
**Table A.10** – Propensity score matching: Balancing test (Health)

	No Health	Health	Diff/se
<b>Geographic Characteristics</b>			
Number of Growing Days (%)	54.301	51.990	2.312 (3.767)
Suitability for Rainfed Crops	5.100	5.286	-0.186 (0.268)
Average Precipitation per Day (over year)	2.918	2.861	0.057 (0.244)
Distance to 2000 city (100km)	2.123	2.511	-0.388 (0.252)
Distance to the Coast (100km)	2.606	2.288	0.318 (0.464)
Malaria Ecology	8.783	8.032	0.751 (1.258)
<b>Historical Characteristics</b>			
Slave Exports, per capita	0.168	0.100	0.069 (0.101)
Railway Contact	0.230	0.170	0.060 (0.056)
Explorer Contact	0.160	0.208	-0.048 (0.054)
Initial Population Density / 100	19.044	19.676	-0.632 (7.381)
Distance to 1400 City (100km)	6.930	6.409	0.520 (0.676)
Distance to 1800 City (100km)	8.565	9.977	-1.412 (1.085)
<b>Mission Characteristics</b>			
Arrival Date	1886.300	1880.861	5.439* (2.430)
Bible Society	0.100	0.236	-0.136* (0.059)
Native Workers	6.260	3.094	3.166 (3.478)
Total Mission Population	335.300	367.160	-31.860 (124.555)
Number of Students	324.130	354.821	-30.691 (123.244)
Physician	0.010	0.255	-0.245*** (0.047)
Observations	206		

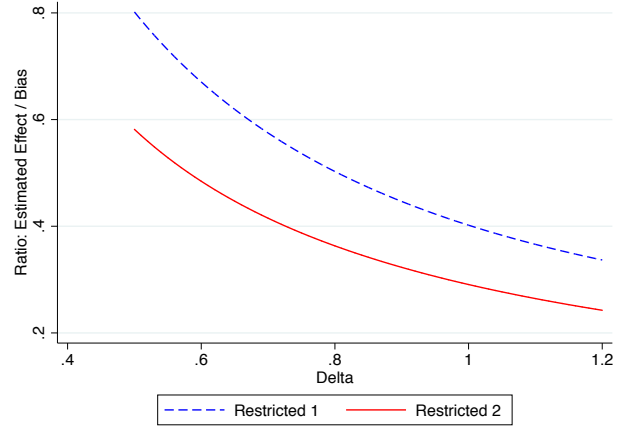
**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table compares geographical and historical characteristics of missions with a health facility and the missions matched using the propensity score. Column 1 presents the results for missions without a health facility. Column 2 presents the results for missions with a health facility. In Column 3 we perform a t-test on the equality of means (robust standard errors in parenthesis). Variables are described in more details in the text.

it would be necessary to know the true value of  $R_{\max}$ . Because there is probably some randomness in the movements of the outcome, it is unlikely that  $R_{\max}$  is equal to one. Oster (2013) uses the R-squared from different randomized experiments as a measure of  $R_{\max}$ . We cannot use the same approach because we analyze historical events. In the literature of the long-term consequences of historical events, in particular of Protestantism (Becker and Woessmann, 2009) and the diffusion of the printing press (Dittmar, 2011) the R-squared rarely exceed 0.65 and are usually close to 0.5. In our regressions, the R-squared never exceed 0.5. Thus, in our computations we choose 0.6, a conservative value of  $R_{\max}$ .

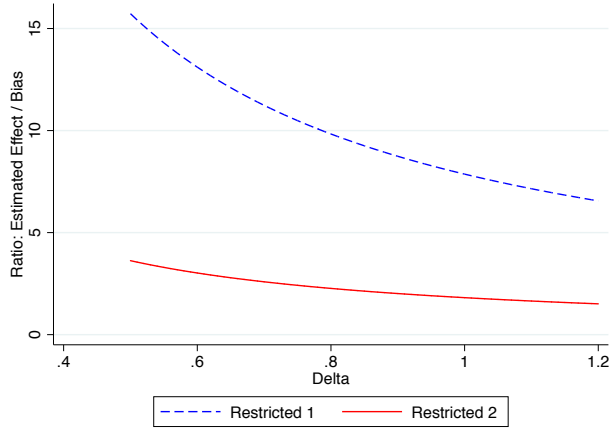
Figure A.6 plots the ratio of the treatment as given in Table ?? over the bias  $B(\delta)$  for different values of  $\delta$  with  $R_{\max} = 0.6$ . In each graph, two different restricted sets of controls  $M$  are chosen. The first set only includes country fixed effects; the second one includes country fixed effects, age, age squared, gender, distance to the closest mission, and distance to the capital city. All the regressions restrict the sample to villages located in 100-km radius of a mission. For newspaper readership and education, the ratio of the treatment over the bias  $B(\delta)$  is higher than one. These results make it unlikely that the entire estimated effect of the distance to the printing press is driven by unobserved variables.



(a) Newspaper Readership



(b) Trust



(c) Education

**Figure A.6** – Magnitude of the Treatment Relative to the Bias for Different Values of  $\delta$ , for  $R2 = 0.6$

**Notes:** Each graph plots the ratio of the treatment over the bias  $B(\delta)$  using two different sets of restricted controls  $M$ . The first set (“Restricted 1”) only includes country fixed effects. The second set (“Restricted 2”) includes country fixed effects, age, age squared, gender, distance to the closest mission, and distance to the capital city. The value  $R_{\max} = 0.6$  is used to compute  $B(\delta)$ .

## A.4 Additional Results

**Table A.11** – Distance to missionary societies and Contemporary outcomes, OLS estimation (100km restriction)

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to society with printing press	-0.015*	0.003	-0.027
	(0.009)	(0.010)	(0.031)
Distance to health	0.002	0.005	0.026
	(0.017)	(0.013)	(0.064)
Distance to school	0.004	-0.004	-0.057
	(0.012)	(0.010)	(0.039)
Observations	28,590	15,511	28,720
Country and Wave FE	Yes	Yes	Yes
Clusters	2,213	917	2,212
R-sq	0.18	0.11	0.24

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

**Table A.12** – Distance to mission investments and Radio and Television, OLS estimation (100km restriction)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Radio	Radio	Radio	Radio	Television	Television	Television	Television
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Distance to printing press	-0.012* (0.006)			-0.013** (0.007)	-0.021* (0.012)			-0.021* (0.012)
Distance to health		-0.005 (0.009)		-0.006 (0.009)		0.008 (0.015)		0.006 (0.015)
Distance to school			-0.006 (0.007)	-0.007 (0.006)			-0.002 (0.012)	-0.003 (0.011)
Observations	28,751	28,751	28,751	28,751	28,661	28,661	28,661	28,661
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	2,213	2,213	2,213	2,213	2,213	2,213	2,213	2,213
R-sq	0.08	0.08	0.08	0.08	0.26	0.25	0.25	0.26

**Notes:** \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

**Table A.13** – Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction), Reporting all the covariates

	(1)	(2)	(3)	(4)
	News	Trust	Education	Clinic
	b/se	b/se	b/se	b/se
Distance to Protestant mission	-0.004 (0.007)	0.001 (0.008)	0.038 (0.031)	0.012 (0.012)
Distance to printing press	-0.024** (0.011)	-0.036*** (0.011)	-0.110*** (0.038)	0.006 (0.021)
Distance to health	-0.003 (0.012)	0.003 (0.013)	0.012 (0.052)	0.040 (0.026)
Distance to school	0.004 (0.009)	-0.007 (0.010)	-0.062** (0.030)	-0.019 (0.016)
Age	-0.003*** (0.001)	0.002 (0.002)	-0.014*** (0.004)	-0.000 (0.001)
Age 2	-0.000* (0.000)	0.000 (0.000)	-0.000*** (0.000)	-0.000 (0.000)
Female	-0.117*** (0.006)	-0.035*** (0.008)	-0.454*** (0.030)	0.001 (0.002)
Distance To City1800	-0.002 (0.012)	0.028*** (0.010)	-0.001 (0.044)	0.008 (0.020)
Initial population density (log)	0.021*** (0.008)	-0.008 (0.008)	0.082*** (0.024)	-0.006 (0.012)
Slave per area (log)	-0.003 (0.012)	-0.001 (0.011)	-0.032 (0.042)	-0.012 (0.023)
Distance Coast	0.009 (0.009)	0.017* (0.009)	0.078** (0.034)	0.013 (0.018)
Growing days in year (%)	0.000*** (0.000)	-0.000 (0.000)	0.001** (0.000)	0.000* (0.000)
Rainfed suitability	0.004 (0.006)	0.001 (0.006)	-0.028 (0.019)	-0.006 (0.010)
Malaria ecology	0.000 (0.002)	0.002 (0.002)	-0.002 (0.006)	0.001 (0.003)
Ruggedness	-0.000*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)	-0.000 (0.000)
Annual Precipitation	-0.000*** (0.000)	0.000 (0.000)	-0.000** (0.000)	-0.000 (0.000)
Distance To Rail	-0.023*** (0.004)	0.002 (0.004)	-0.049*** (0.016)	-0.003 (0.007)
Distance To Explorer	-0.009 (0.009)	0.009 (0.007)	0.005 (0.027)	-0.028** (0.012)
Closest Mission Size	0.000*** (0.000)	-0.000* (0.000)	0.000*** (0.000)	0.000 (0.000)
Closest Mission Has Education	-0.024 (0.022)	0.033 (0.022)	-0.196** (0.077)	0.030 (0.038)
Closest Mission Hospital	0.043 (0.036)	0.031 (0.043)	0.172 (0.112)	-0.035 (0.053)
Closest Mission Has Health	-0.031 (0.029)	0.018 (0.031)	0.030 (0.089)	0.026 (0.048)
Distance To Kgarten	0.004 (0.012)	-0.009 (0.012)	0.037 (0.040)	0.031* (0.019)
Distance to Catholic mission	0.004 (0.009)	-0.008 (0.009)	0.079* (0.043)	-0.015 (0.017)
Observations	28,590	15,511	28,720	28,267
Country and Wave FE	Yes	Yes	Yes	Yes
Clusters	2,213	917	2,212	2,175
R-sq	0.18	0.11	0.24	0.14

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates with all the covariates. The unit of observation is an individual. Dependent variable are contemporary outcomes described in more details in the text. Controls are the individual-, village-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

APPENDIX A. APPENDIX TO PART 1

**Table A.14** – Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction), Standardized coefficients (Beta Coefficients)

	(1) News beta/se	(2) Trust beta/se	(3) Education beta/se	(4) Clinic beta/se
Distance to Protestant mission	-0.007 (0.007)	0.001 (0.008)	0.020 (0.031)	0.025 (0.012)
Distance to printing press	-0.053** (0.011)	-0.081*** (0.011)	-0.063*** (0.038)	0.013 (0.021)
Distance to health	-0.007 (0.012)	0.006 (0.013)	0.008 (0.052)	0.090 (0.026)
Distance to school	0.010 (0.009)	-0.016 (0.010)	-0.040** (0.030)	-0.045 (0.016)
Age	-0.087*** (0.001)	0.070 (0.002)	-0.110*** (0.004)	-0.004 (0.001)
Age 2	-0.058* (0.000)	0.017 (0.000)	-0.180*** (0.000)	-0.006 (0.000)
Female	-0.121*** (0.006)	-0.036*** (0.008)	-0.120*** (0.030)	0.001 (0.002)
Distance To City1800	-0.006 (0.012)	0.078*** (0.010)	-0.000 (0.044)	0.021 (0.020)
Initial population density (log)	0.072*** (0.008)	-0.026 (0.008)	0.071*** (0.024)	-0.019 (0.012)
Slave per area (log)	-0.006 (0.012)	-0.002 (0.011)	-0.018 (0.042)	-0.025 (0.023)
Distance Coast	0.033 (0.009)	0.062* (0.009)	0.074** (0.034)	0.046 (0.018)
Growing days in year (%)	0.352*** (0.000)	-0.092 (0.000)	0.247** (0.000)	0.292* (0.000)
Rainfed suitability	0.015 (0.006)	0.003 (0.006)	-0.025 (0.019)	-0.022 (0.010)
Malaria ecology	0.006 (0.002)	0.031 (0.002)	-0.010 (0.006)	0.023 (0.003)
Ruggedness	-0.080*** (0.000)	-0.033 (0.000)	-0.128*** (0.000)	-0.077 (0.000)
Annual Precipitation	-0.378*** (0.000)	0.114 (0.000)	-0.262** (0.000)	-0.284 (0.000)
Distance To Rail	-0.095*** (0.004)	0.010 (0.004)	-0.051*** (0.016)	-0.012 (0.007)
Distance To Explorer	-0.027 (0.009)	0.028 (0.007)	0.004 (0.027)	-0.084** (0.012)
Closest Mission Size	0.060*** (0.000)	-0.037** (0.000)	0.057*** (0.000)	0.006 (0.000)
Closest Mission Has Education	-0.021 (0.022)	0.027 (0.022)	-0.043** (0.077)	0.024 (0.038)
Closest Mission Hospital	0.022 (0.036)	0.014 (0.043)	0.023 (0.112)	-0.017 (0.053)
Closest Mission Has Health	-0.023 (0.029)	0.012 (0.031)	0.006 (0.089)	0.018 (0.048)
Distance To Kgarten	0.011 (0.012)	-0.023 (0.012)	0.027 (0.040)	0.083* (0.019)
Distance to Catholic mission	0.008 (0.009)	-0.017 (0.009)	0.039* (0.043)	-0.028 (0.017)
Observations	28,590	15,511	28,720	28,267
Country and Wave FE	Yes	Yes	Yes	Yes
Clusters	2,213	917	2,212	2,175
R-sq	0.18	0.11	0.24	0.14

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates with all the covariates. Coefficients are beta coefficients. The unit of observation is an individual. Dependent variable are contemporary outcomes described in more details in the text. Controls are the individual-, village-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.



## A.5 Robustness Checks

### A.5.1 Distance

**Table A.15** – Distance to printing presses and Contemporary outcomes, OLS estimation, No restriction on distance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	News	News	News	Trust	Trust	Trust	Trust	Trust	Education	Education	Education	Education	Clinic	Clinic	Clinic	Clinic
	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se	b/se
Distance to printing press	-0.014 (0.010)			-0.024** (0.011)	-0.057*** (0.011)			-0.036*** (0.011)	0.001 (0.035)			-0.110*** (0.038)	0.023 (0.018)			0.006 (0.021)
Distance to health		-0.017 (0.011)		-0.003 (0.012)		0.023* (0.012)		0.003 (0.013)		-0.072 (0.046)		0.012 (0.052)	0.017 (0.023)			0.040 (0.026)
Distance to school			0.009 (0.009)	0.004 (0.009)			-0.001 (0.011)	-0.007 (0.010)			-0.070** (0.029)	-0.062** (0.030)		-0.011 (0.016)		-0.019 (0.016)
Observations	45,825	45,825	45,825	28,590	24,863	24,863	24,863	15,511	46,116	46,116	46,116	28,720	45,371	45,371	45,371	28,267
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	3,621	3,621	3,621	2,213	1,574	1,574	1,574	917	3,620	3,620	3,620	2,212	3,563	3,563	3,563	2,175
R-sq	0.18	0.18	0.18	0.18	0.10	0.10	0.10	0.11	0.28	0.28	0.28	0.24	0.11	0.10	0.10	0.14

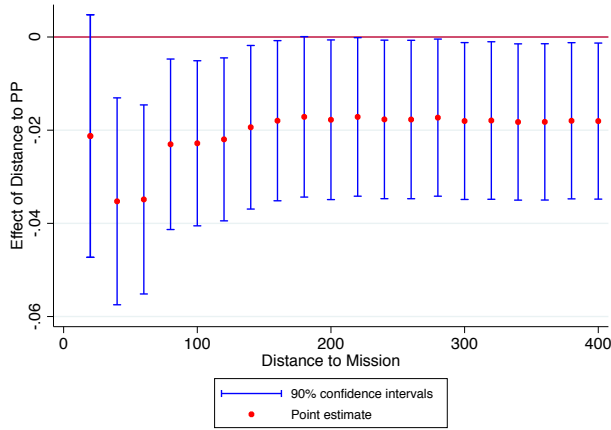
**Notes:** \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects.

**Table A.16** – Distance to printing presses and Contemporary Outcomes, OLS estimation using a binary variable (100km restriction)

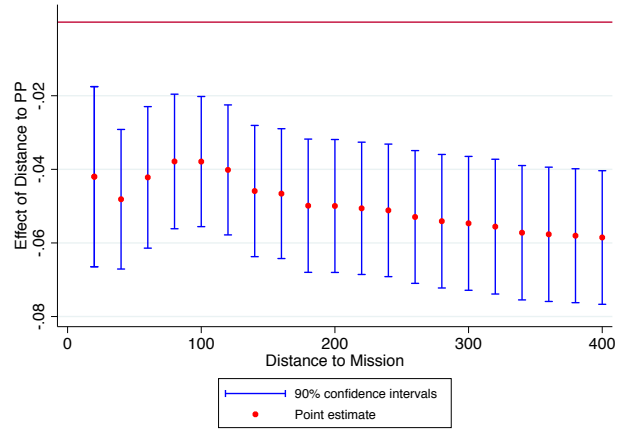
	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
D(<25km Protestant mission with printing press)	0.100** (0.044)	0.075* (0.044)	0.559*** (0.134)
Observations	28,590	15,511	28,720
Country and Wave FE	Yes	Yes	Yes
Clusters	2,213	917	2,212
R-sq	0.18	0.11	0.24

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary outcomes described in more details in the text. The explanatory variable is a binary variable equal to 1 if the individual lives within a 25km radius around a mission with a printing press. Standard errors in parentheses are clustered at the town-level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

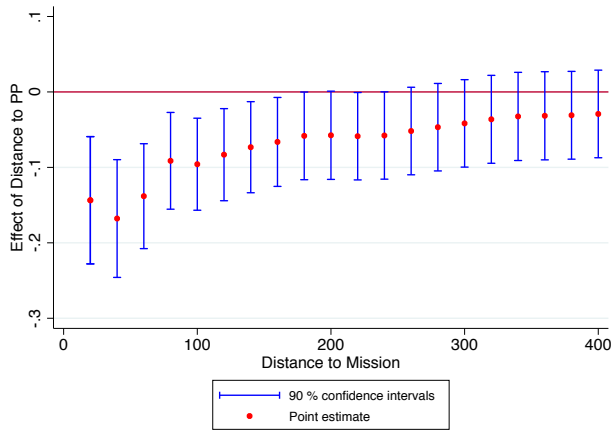
## A.5.2 Additional Robustness Checks



(a) Newspaper Readership



(b) Trust



(c) Education

**Figure A.7** – Effect of Distance to the Printing Press on Contemporary Outcomes, for Different Restrictions on Proximity to Missions

**Table A.17** – Printing Press and Contemporary Outcomes, Probit Estimation

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to printing press	-0.070** (0.031)	-0.106*** (0.030)	-0.031 (0.035)
Distance to health	-0.006 (0.037)	0.009 (0.039)	0.001 (0.040)
Distance to school	0.008 (0.028)	-0.020 (0.029)	-0.066** (0.029)
Observations	28590	15511	28720
Country and Wave FE	Yes	Yes	Yes
Clusters	2213	917	2212
PseudoR2	0.148	0.0870	0.190

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table reports Probit estimates. The unit of observation is an individual. The dependent variable is newspaper readership nowadays. Standard errors in parentheses are clustered by village. The controls are the individual-, village- and ethnicity-level controls described in the text and the distance to the closest mission without a printing press. All specifications include country fixed effects

**Table A.18** – Printing Press and Contemporary Outcomes, with Potential “Bad Controls”

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to printing press	-0.022** (0.010)	-0.038*** (0.011)	-0.097** (0.038)
Distance to health	-0.001 (0.011)	0.001 (0.013)	0.015 (0.045)
Distance to school	0.009 (0.009)	-0.007 (0.011)	-0.056** (0.029)
Observations	28446	15430	28570
Country and Wave FE	Yes	Yes	Yes
Clusters	2213	917	2212
R2	0.214	0.112	0.275
F-Statistic	96.78	27.13	105.8

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table reports OLS estimates. The unit of observation is an individual. Standard errors in parentheses are clustered by village. The specification here includes “bad controls” (that is contemporary possible outcomes: religion, water constraints, cash constraints, and income proxied with light density). All specifications include country fixed effects.

**Table A.19** – Distance to the Largest Missions and Contemporary Outcomes

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to Big Mission	-0.003 (0.013)	0.024* (0.013)	0.049 (0.049)
Distance to school	0.006 (0.009)	-0.007 (0.011)	-0.058* (0.031)
Distance to health	-0.000 (0.013)	-0.000 (0.014)	0.012 (0.055)
Observations	28590	15511	28720
Country and Wave FE	Yes	Yes	Yes
Clusters	2213	917	2212
R2	0.182	0.110	0.241
F-Statistic	73.88	29.27	91.14

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table reports OLS estimates. The unit of observation is an individual. Standard errors in parentheses are clustered by village. The baseline controls are the individual-, village- and ethnicity-level controls described in the text and the distance to the closest mission without a printing press. All specifications include country fixed effects.

**Table A.20** – Printing Press and Contemporary Outcomes, Clustering at the Mission Level

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to printing press	-0.024* (0.013)	-0.036*** (0.010)	-0.110*** (0.043)
Distance to health	-0.003 (0.015)	0.003 (0.012)	0.012 (0.057)
Distance to school	0.004 (0.010)	-0.007 (0.011)	-0.062** (0.031)
Observations	28590	15511	28720
Country and Wave FE	Yes	Yes	Yes
Clusters	431	343	431
R2	0.182	0.111	0.241
F-Statistic	62.19	43.78	70.05

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table reports OLS estimates. The unit of observation is an individual. The dependent variable is newspaper readership. Standard errors in parentheses are clustered by closest mission. All specifications include country fixed effects.

**Table A.21** – Printing Press and Newspaper Readership, Former British Colonies

	(1)	(2)	(3)
	News	Trust	Education
	b/se	b/se	b/se
Distance to printing press	-0.054*** (0.016)	-0.058*** (0.019)	-0.162** (0.069)
Distance to health	-0.008 (0.023)	-0.019 (0.025)	-0.005 (0.086)
Distance to school	0.004 (0.016)	-0.002 (0.018)	-0.032 (0.057)
Observations	11680	6234	11731
Country and Wave FE	Yes	Yes	Yes
Clusters	1103	467	1102
R2	0.218	0.0661	0.223
F-Statistic	59.83	9.055	69.36

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The table reports OLS estimates. The unit of observation is an individual. The dependent variable is newspaper readership. The sample of countries is reduced to former British colonies. Standard errors in parentheses are clustered by village. All specifications include country fixed effects.

**Table A.22** – Proximity to a printing press and Contemporary outcomes: Additional Robustness with one-way clustering

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	News b/se	News b/se	News b/se	News b/se	Trust b/se	Trust b/se	Trust b/se	Trust b/se	Education b/se	Education b/se	Education b/se	Education b/se
Distance to printing press	-0.024 (0.015)	-0.059*** (0.014)	0.021 (0.026) -0.078** (0.030)	0.022 (0.020) -0.078*** (0.024)	-0.036*** (0.006)	-0.038*** (0.009)	-0.027*** (0.010) -0.015 (0.016)	-0.027* (0.016) -0.015 (0.021)	-0.105** (0.045)	-0.130** (0.056)	-0.059 (0.080) -0.079 (0.098)	-0.064 (0.063) -0.080 (0.079)
Distance to health				-0.004 (0.012)			0.002 (0.013)	0.002 (0.013)				0.011 (0.052)
Distance to school				0.003 (0.009)			-0.007 (0.010)	-0.007 (0.010)				-0.062** (0.030)
Observations	28,590	28,590	28,590	28,590	15,511	15,511	15,511	15,511	28,720	28,720	28,720	28,720
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	52	52	52	2,213	48	48	48	917	52	52	52	2,212
R-sq	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.24	0.24	0.24	0.24

**Notes:** \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary measures of political participation described in more details in the text. Standard errors in parentheses are clustered at the closest mission with a printing press and a publication record level. Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.



**Table A.23** – Proximity to a printing press and Contemporary outcomes: Additional Robustness with two-way clustering

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance to printing press	News b/se -0.024 (0.015)	News b/se -0.059*** (0.014)	News b/se 0.021 (0.026)	News b/se 0.022 (0.020)	Trust b/se -0.036*** (0.006)	Trust b/se -0.038*** (0.009)	Trust b/se -0.027*** (0.010)	Trust b/se -0.027* (0.016)	Education b/se -0.105** (0.045)	Education b/se -0.130** (0.056)	Education b/se -0.059 (0.080)	Education b/se -0.064 (0.063)
Distance publication record												
Distance to health												
Distance to school												
Observations	28,590	28,590	28,590	28,590	15,511	15,511	15,511	15,511	28,720	28,720	28,720	28,720
Country and Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	52	52	52	2,213	48	48	48	917	52	52	52	2,212
R-sq	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.24	0.24	0.24	0.24

**Notes:** \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. The Table reports OLS estimates. The unit of observation is an individual. Dependent variables are contemporary measures of political participation described in more details in the text. Standard errors in parentheses are clustered two ways (at the closest mission with a printing press and a publication record level and at the village level). Controls are the individual-, town-, and ethnicity-level controls described in the text. All specifications include country and Afrobarometer round fixed effects. The sample is restricted to individuals living in a 100km radius around a historical missionary settlement.

# Appendix B

## Appendix to Part 2

### B.1 Data

#### B.1.1 American Community Survey

The American Community Survey reports ancestry for each respondent. Each individual is given the possibility to report several ancestries, on the three years of our sample (2008-2010), around 2/3 of the respondents reported a single ancestry and 1/3 reported two. The raw data has over 1000 categories reported, but these are not the finally published categories. The self-reported ancestries are aggregate into a shorter list of categories (see below). The aggregation is “the result of Census Bureau’s own research and consultations with many ethnic experts.” (ACS, 2010)

The ancestry question allows respondents to report one or more ancestry groups. The variable definition report explains in more details how they handle more complex cases of ancestry reporting (ACS, 2010)

“The ancestry question allowed respondents to report one or more ancestry groups. Generally, only the first two responses reported were coded. If a response was in terms of a dual ancestry, for example, “Irish English,” the person was assigned two codes, in this case one for Irish and another for English. However, in certain cases, multiple responses such as “French Canadian,” “Scotch-Irish,” “Greek Cypriot,” and “Black Dutch” were assigned a single code reflecting their

status as unique groups. If a person reported one of these unique groups in addition to another group, for example, “Scotch-Irish English,” resulting in three terms, that person received one code for the unique group (Scotch-Irish) and another one for the remaining group (English). If a person reported “English Irish French,” only English and Irish were coded. If there were more than two ancestries listed and one of the ancestries was a part of another, such as “German Bavarian Hawaiian,” the responses were coded using the more detailed groups (Bavarian and Hawaiian).

The Census Bureau accepted “American” as a unique ethnicity if it was given alone or with one other ancestry. There were some groups such as “American Indian,” “Mexican American,” and “African American” that were coded and identified separately.”

The final list of possible ancestries is as following:

Alsatian, Austrian, Basque, Belgian, Flemish, British, British Isles, Danish, Dutch, English, Finnish, French, German, Prussian, Greek, Icelander, Irish, Italian, Sicilian, Luxembourg, Maltese, Norwegian, Portuguese, Scotch Irish, Scottish, Swedish, Swiss, Welsh, Scandinavian, Celtic, Albanian, Belorussian, Bulgarian, Croatian, Czech, Bohemian, Czechoslovakian, Estonian, German Russian, Rom, Hungarian, Latvian, Lithuanian, Macedonian, Polish, Romanian, Russian, Serbian, Slovak, Slovene, Georgia CIS, Ukrainian, Yugoslavian, Herzegovinian, Slavic, Slavonian, Northern European, Western European, Eastern European, European, Spaniard, Mexican, Mexican American, Mexicano, Chicano, Mexican American Indian, Mexican State, Costa Rican, Guatemalan, Honduran, Nicaraguan, Panamanian, Salvadoran, Central American, Argentinean, Bolivian, Chilean, Colombian, Ecuadorian, Paraguayan, Peruvian, Uruguayan, Venezuelan, South American, Latin American, Latin, Latino, Puerto Rican, Cuban, Dominican, Hispanic, Spanish, Spanish American, Bahamian, Barbadian, Belizean, Jamaican, Dutch West Indian, Trinidadian Tobagonian, British West Indian, Antigua and Barbuda, Grenadian, Vincent-Grenadine Islander, St Lucia Islander, West Indian, Haitian, Other West Indian, Brazilian, Guyanese, Algerian, Egyptian, Moroccan, Iranian, Iraqi, Israeli, Jordanian, Lebanese, Syrian, Armenian, Turkish, Yemeni, Kurdish, Palestinian, Assyrian, Chaldean, Mideast, Arab, Arabic, Other Arab, Cameroon, Cape Verdean, Ethiopian, Eritrean, Ghanaian, Kenyan, Liberian, Nige-

rian, Senegalese, Sierra Leonean, Somalian, South African, Sudanese, Other Subsaharan African, Western African, African, Afghan, Bangladeshi, Nepali, Asian Indian, Bengali, East Indian, Punjab, Pakistani, Sri Lankan, Burmese, Cambodian, Chinese, Cantonese, Mongolian, Filipino, Indonesian, Japanese, Okinawan, Korean, Laotian, Hmong, Malaysian, Thai, Taiwanese, Vietnamese, Eurasian, Amerasian, Asian, Other Asian, Australian, New Zealander, Polynesian, Hawaiian, Samoan, Tongan, Micronesian, Guamanian, Chamorro Islander, Fijian, Pacific Islander, Other Pacific, Afro American, Afro, African American, Black, Negro, Creole, Central American Indian, South American Indian, Native American, Indian, Cherokee, American Indian, Aleut, Eskimo, White, Anglo, Appalachian, Pennsylvania German, Canadian, French Canadian, Acadian, Cajun, American or United States, Texas, North American, Mixture, Uncodable entries, Other groups, Other responses, and Not reported.

## B.1.2 Cultural Distance

### Notes on the Computation of the Fixation Index

**Heterozygosity** The heterozygosity of a given marker measures the extent of the allele diversity in a given population. More exactly, it measures the probability that two markers drawn at random from the population have the same allele.

**Fixation Index  $F_{ST}$**  Let  $F_{ST}$  be the fixation index. In a given population, consider there are  $M$  alleles with average frequency  $\bar{x}_m$ ,  $m \in \{1, \dots, M\}$  for each allele. There are  $K$  subgroups in the population with frequencies  $x_m^k$ ,  $k \in \{1, \dots, K\}$  for each allele. For this population, we can define:

$$H_T = \text{Expected Overall Heterozygosity} = 1 - \sum_{m=1}^M \bar{x}_m^2$$

$$H_s = \text{Expected Heterozygosity across Subpopulations} = \frac{1}{K} \sum_{k=1}^K \left[ 1 - \sum_{m=1}^M (x_m^k)^2 \right]$$

The  $F_{ST}$  is then defined by:

	(1) b/se
WVS Perceptions of Life	39.136*** (0.413)
WVS Work	-79.163*** (1.004)
WVS Family	-27.299*** (1.578)
WVS Politics and Society	0.353 (0.572)
WVS Religion and Morale	49.969*** (0.610)
Religion	1617.339*** (4.253)
Language	-1121.412*** (6.187)
Observations	42,036
R2	.98900206
F Stat.	540020.15

**Table B.1** –  $F_{ST}$  and Other Measures of Cultural Distance in the ACS dataset

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an ancestry at a given location. The dependent variable is the Individual-Local  $F_{ST}^{aP}$ . Standard errors in parentheses are clustered at the PUMA level. The explanatory variables are alternative measures of cultural distance computed by Spolaore and Wacziarg (2015)

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

If members of the population interbreed freely and as much within the groups than between them, we expect  $H_T = H_S$  and  $F_{ST} = 0$ . There is free exchange of genetic material, and hence the genetic distance will on average be equal to 0. On the opposite case, suppose that individuals only interbreed between members of their own group. In this case there is full inbreeding for each group. It is an established result that heterozygosity converges to 0 when there is full inbreeding. Therefore, in the limiting case of no exchange between groups the genetic distance is maximal,  $F_{ST} = 1$ .

Genetic Distance Data

## B.2 Additional Results

### B.2.1 Selection on Unobservables

Suppose there is a set of unobservable explanatory variables  $W'$ . Since these variables are unobserved, they are not included in regression (4.4). The *proportional selection assumption* (PSA) states that  $\delta C_{WX} = \frac{C_{W'X}}{V_{W'}}$ .  $X$  is the treatment variable,  $W$  is the set of observed covariates,  $C_{W'X}$  is the covariance of  $W'$  and  $X$  and  $V_{W'}$  is the variance of  $W'$ .  $\delta$  is a measure of the relationship between  $C_{WX}$  and  $C_{W'X}$ . The PSA assumption is key in the approach as it states that the relationship between the observed covariates  $W$  and the treatment  $X$  is informative about the relationship between the unobserved covariates  $W'$  and  $X$ , from which the bias is coming.

Consider the following three regressions:

$$y = \gamma X + W + W' + \varepsilon_{\max} \tag{M-max}$$

$$y = \xi X + M + \varepsilon_1 \tag{M-1}$$

$$y = \Lambda X + W + \varepsilon_2 \tag{M-2}$$

Let  $R_{\max}$  be the R-squared of the full model regression (M-max).  $R_2$  is the R-squared of the regression (M-2) including all the observed covariates.  $R_1$  is the R-squared of regression (M-1) including only a restricted set of covariates  $M$ .  $M$  is a set of observed controls that do not have a related unobserved component and are orthogonal to  $W$  and  $W'$  (Oster, 2013).

According to Oster (2013), under the PSA and when  $\delta$  is close to one,  $B(\delta) = \delta \frac{(\xi - \Lambda)(R_{\max} - R_2)}{R_2 - R_1}$  is (i) equal to the unobserved bias if  $\delta = 1$ ; (ii) a close upper bound on the bias if  $\delta < 1$ ; (iii) and a close lower bound on the bias if  $\delta > 1$ .<sup>1</sup>

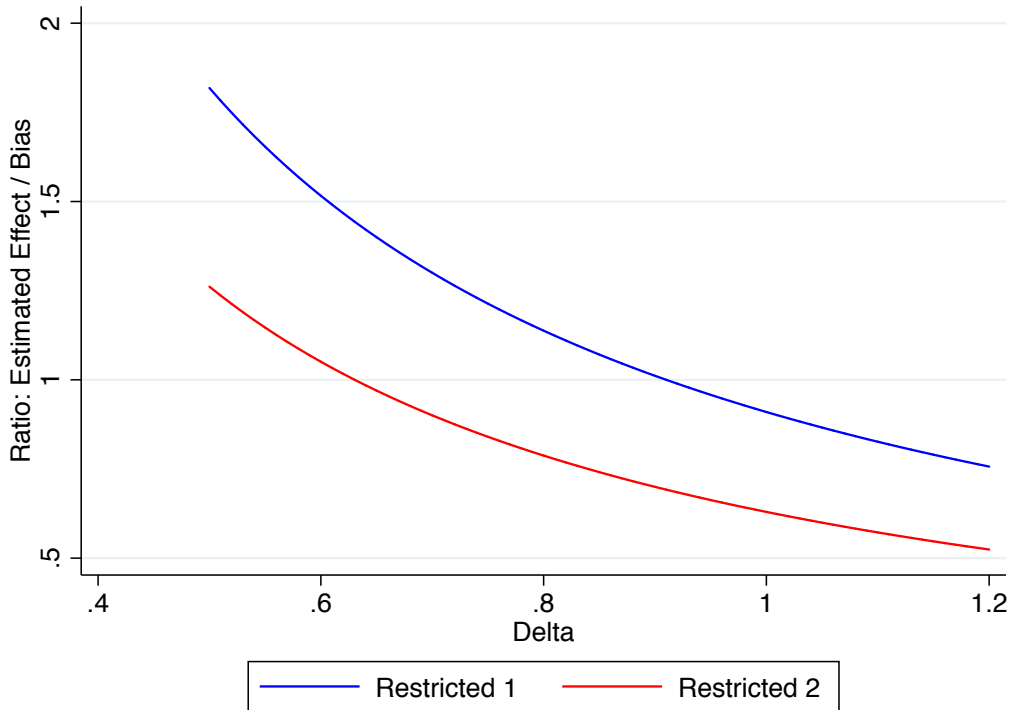
We can compute the bias due to unobserved variables from the movements in the treatment effect due to the added controls using the ratio  $B(\delta)$ . However, to compute the ratio

---

<sup>1</sup>If  $W$  is selected randomly from  $\{W, W'\}$ , then  $\delta = 1$ . If  $W$  is the most important set of controls from  $\{W, W'\}$  then  $\delta < 1$ .

it would be necessary to know the true value of  $R_{\max}$ . Oster (2013) uses the R-squared from different randomized experiments as a measure of  $R_{\max}$ . We cannot use the same approach because we are dealing with non-randomizable events (migration). Ager and Brukner (2013)'s regression have an R2 of around 0.5, so this is the value we use.

Figure B.1 plots the ratio of the treatment as given in the Column (4) of Table 4.3 over the bias  $B(\delta)$  for different values of  $\delta$  with  $R_{\max} = 0.5$ . Two different restricted sets of controls  $M$  are chosen. The first set only includes country fixed effects (ancestry and state); the second one includes country fixed effects, age, age squared, and sex. Our results suggest that for the bias unobservables to fully explain the effect of the  $F_{ST}$ , it means that selection on unobservables represent from 50% to 150% of selection on observables, which is unlikely.



**Figure B.1** – Magnitude of the Treatment Relative to the Bias for different Values of  $\delta$ , For  $R2 = 0.5$

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The plot shows the ratio of the treatment over the bias  $B(\delta)$  using two different sets of restricted controls  $M$ . The first set (“Restricted 1”) only includes state and ancestry fixed effects. The second set (“Restricted 2”) includes state and ancestry fixed effects, age, age squared, and sex. The value  $R_{\max} = 0.5$  is used to compute  $B(\delta)$

**Table B.2** – Effect of Religious Distance and Ancestry-Inherited Regional Characteristics on Income

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
Religious	0.231*** (0.029)	0.006 (0.042)	0.276*** (0.046)	0.198*** (0.064)
<b>Fractionalization</b>				
High Region ethnic	0.117*** (0.025)			0.067** (0.027)
Religious x High Region ethnic	-0.211*** (0.050)			-0.132** (0.055)
<b>Trust</b>				
High Trust		-0.131*** (0.021)		-0.099*** (0.025)
Religious x High Trust		0.201*** (0.041)		0.137*** (0.047)
<b>Xenophobia</b>				
High Xenophobia			0.046** (0.022)	0.067*** (0.021)
Religious x High Xenophobia			-0.136*** (0.045)	-0.168*** (0.042)
Observations	871,470	871,470	871,470	871,470
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

## B.2.2 Alternative Measures of Cultural Distance



**Table B.3** – Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - General Perceptions of Life Measure

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
WVS a	-0.117*** (0.010)	-0.121*** (0.010)	-0.119*** (0.009)	-0.111*** (0.013)
<b>Fractionalization</b>				
High Region ethnic	0.027 (0.027)			0.010 (0.028)
WVS a x High Region ethnic	-0.012 (0.014)			-0.004 (0.014)
<b>Trust</b>				
High Trust		-0.080*** (0.026)		-0.092*** (0.029)
WVS a x High Trust		0.030** (0.013)		0.036** (0.016)
<b>Xenophobia</b>				
High Xenophobia			-0.002 (0.033)	0.013 (0.035)
WVS a x High Xenophobia			-0.004 (0.017)	-0.012 (0.018)
Observations	879,054	879,054	879,054	879,054
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

**Table B.4** – Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Work-Ethic Measure

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
WVS c	-0.040*** (0.014)	-0.080*** (0.010)	-0.085*** (0.010)	-0.052*** (0.017)
<b>Fractionalization</b>				
High Region ethnic	0.074*** (0.016)			0.041** (0.017)
WVS c x High Region ethnic	-0.052*** (0.016)			-0.029* (0.017)
<b>Trust</b>				
High Trust		-0.094*** (0.014)		-0.062*** (0.016)
WVS c x High Trust		0.062*** (0.014)		0.034** (0.016)
<b>Xenophobia</b>				
High Xenophobia			-0.068*** (0.015)	-0.030* (0.015)
WVS c x High Xenophobia			0.047*** (0.015)	0.012 (0.015)
Observations	879,054	879,054	879,054	879,054
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

**Table B.5** – Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Family Values Measure

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
WVS d	-0.139*** (0.017)	-0.208*** (0.016)	-0.190*** (0.012)	-0.172*** (0.022)
<b>Fractionalization</b>				
High Region ethnic	0.083*** (0.021)			0.055*** (0.021)
WVS d x High Region ethnic	-0.065*** (0.021)			-0.041** (0.021)
<b>Trust</b>				
High Trust		-0.110*** (0.018)		-0.069*** (0.019)
WVS d x High Trust		0.099*** (0.018)		0.062*** (0.020)
<b>Xenophobia</b>				
High Xenophobia			-0.081*** (0.020)	-0.060*** (0.021)
WVS d x High Xenophobia			0.061*** (0.020)	0.044** (0.020)
Observations	879,054	879,054	879,054	879,054
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

**Table B.6** – Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Religion and Morale Measure

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
WVS f	-0.044*** (0.015)	0.043*** (0.016)	0.047*** (0.012)	-0.025 (0.020)
<b>Fractionalization</b>				
High Region ethnic	-0.135*** (0.020)			-0.112*** (0.020)
WVS f x High Region ethnic	0.133*** (0.016)			0.105*** (0.016)
<b>Trust</b>				
High Trust		0.019 (0.021)		-0.018 (0.020)
WVS f x High Trust		-0.051*** (0.016)		-0.006 (0.017)
<b>Xenophobia</b>				
High Xenophobia			0.061*** (0.019)	0.048** (0.019)
WVS f x High Xenophobia			-0.073*** (0.015)	-0.057*** (0.015)
Observations	879,054	879,054	879,054	879,054
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

**Table B.7** – Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - All Categories Measure

	(1)	(2)	(3)	(4)
	b/se	b/se	b/se	b/se
WVS	-0.090*** (0.010)	-0.101*** (0.013)	-0.094*** (0.010)	-0.082*** (0.012)
<b>Fractionalization</b>				
High Region ethnic	0.120*** (0.045)			0.066 (0.047)
WVS x High Region ethnic	-0.038** (0.016)			-0.020 (0.017)
<b>Trust</b>				
High Trust		-0.107*** (0.037)		-0.121*** (0.036)
WVS x High Trust		0.027** (0.013)		0.033** (0.013)
<b>Xenophobia</b>				
High Xenophobia			0.089* (0.049)	0.057 (0.052)
WVS x High Xenophobia			-0.038** (0.017)	-0.026 (0.018)
Observations	879,054	879,054	879,054	879,054
State FE	Yes	Yes	Yes	Yes
AncestryFE	Yes	Yes	Yes	Yes
Clusters	2,069	2,069	2,069	2,069
R2	0.25	0.25	0.25	0.25

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

### B.2.3 Polarization

**Table B.8** – Effect of  $F_{ST}$  on income - Controlling for Average Polarization

	(1)	(2)	(3)
	All	Non Migrants	Migrants
	b/se	b/se	b/se
Log Fst	0.069*** (0.016)	0.084*** (0.019)	-0.024 (0.036)
Local Pol	0.000 (0.000)	-0.000 (0.000)	0.000*** (0.000)
Local Fract	-0.112*** (0.019)	-0.082*** (0.020)	-0.120*** (0.034)
Observations	906,729	729,861	176,868
State FE	Yes	Yes	Yes
Ancestry FE	Yes	Yes	Yes
Clusters	2,069	2,069	2,068
R2	0.25	0.26	0.26

**Notes:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The Table reports OLS estimates. The unit of observation is an individual. The dependent variable is the log of the individual's earnings. Standard errors in parentheses are clustered at the PUMA level. Controls are the individual level controls described in the text. All specifications include fixed US-state and continent ancestry fixed effects.

# List of Figures

1	The World Missionary Conference in Edinburgh, 1910 . . . . .	30
1.1	Mission Stations With and Without a Printing Press in 1903 . . . . .	49
1.2	Effect of Distance to the Printing Press on Contemporary Outcomes. Locally Weighted Scatterplot Smoothing . . . . .	63
1.3	Voronoi diagram of Africa . . . . .	68
2.1	David Livingstone’s Medical Oath . . . . .	80
2.2	HIV prevalence rates by country . . . . .	106
2.3	HIV prevalence estimates at the subnational level . . . . .	107
2.4	DHS locations . . . . .	108
2.5	Missions and missionary investment locations . . . . .	109
2.6	Road Density . . . . .	110
3.1	An illustration of common descent . . . . .	117
3.2	A model of gene flow . . . . .	122
3.3	$F_{ST}$ between 42 populations, from Cavalli-Sforza (1994) . . . . .	126
3.4	Correlation between genetic distance ( $F_{ST}$ ) and cultural distance (from the WVS) . . . . .	129
3.5	Correlation between $F_{ST}$ and predicted $F_{ST}$ . . . . .	130
3.6	Multidimensional scaling of the $F_{ST}$ . . . . .	131
4.1	Average Hourly Income in PUMA region . . . . .	141
4.2	Average $F_{ST}$ in PUMA region . . . . .	143



4.3	Correlation between the $F_{ST}$ and its predicted value from other measures of cultural distance . . . . .	144
4.4	Marginal Effect of Fst on Income For Different Ancestries . . . . .	153
4.5	Marginal Effect of Fst on Income as a Function of Time Spent in the USA, First-Generation Migrants Only . . . . .	155
4.6	Ancestry Segregation and Ancestry-inherited levels of Trust, at the PUMA area	159
4.7	Ancestry Segregation and Ancestry-inherited levels of Xenophobia, at the PUMA area . . . . .	159
4.8	Ancestry Segregation and Ancestry-inherited levels of Ethnic Fractionalization, at the PUMA area . . . . .	160
A.1	Part of the Statistical index from the <i>Geography Atlas of Protestant missions</i> , 1903 . . .	172
A.2	Plate Representing all the Missionary Stations in 1903, West Africa ( <i>Geography Atlas of Protestant missions</i> , 1903) . . . . .	179
A.3	Distance to the Printing Press and Number of Newspapers . . . . .	181
A.4	Distance to the Printing Press and Publication Date of the First Newspaper	181
A.5	Distribution of the distance of the Afrobarometer towns to the closest mission	183
A.6	Magnitude of the Treatment Relative to the Bias for Different Values of $\delta$ , for $R^2 = 0.6$ . . . . .	193
A.7	Effect of Distance to the Printing Press on Contemporary Outcomes, for Different Restrictions on Proximity to Missions . . . . .	201
B.1	Magnitude of the Treatment Relative to the Bias for different Values of $\delta$ , For $R^2 = 0.5$ . . . . .	212

# List of Tables

1.1	Determinants of the location of missions (Comparing towns close and towns far from Missions) . . . . .	54
1.2	Determinants of the location of the printing press (Comparing missions with and without the printing press) . . . . .	55
1.3	Characteristics of and investments performed by missions with and without a printing press . . . . .	56
1.4	Distance to a mission and Contemporary outcomes, OLS estimation (100km restriction) . . . . .	60
1.5	Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction) . . . . .	62
1.6	Distance to mission investments and Political Participation, Considering only democracies, OLS estimation (100km restriction) . . . . .	65
1.7	Distance to mission investments and Contemporary outcomes, Matching estimation . . . . .	69
1.8	Proximity to a printing Press and Contemporary outcomes: Effect of the regional development of the publishing industry (1923), OLS estimation (100km restriction) . . . . .	77
2.1	Summary Statistics: Characteristics of Missions with Health Investments . .	92
2.2	Summary Statistics: Characteristics of Missions with the Printing Press Investments . . . . .	93
2.3	Missions and HIV . . . . .	95
2.4	Missionary Investments and HIV . . . . .	99

2.5	Missionary Investments and HIV, Matching Approach . . . . .	101
2.6	Robustness Checks . . . . .	105
4.1	Summary Statistics ACS . . . . .	147
4.2	Summary Statistics of the Diversity Measures . . . . .	148
4.3	Baseline Effect of Fst on Income . . . . .	150
4.4	Effect of Fst and Local Ancestry Divisions on Income . . . . .	158
4.5	Effect of Fst and Ancestry-Inherited Individual Characteristics on Income . .	163
4.6	Effect of Fst and Ancestry-Inherited Regional Characteristics on Income . .	165
4.7	Effect of Fst and Ancestry-Inherited Regional Characteristics on the Probability of Being Self Employed and the Income of the Self-Employed . . . . .	167
A.1	Variables from Afrobarometer, Description . . . . .	180
A.3	Summary statistics: Distance of the Afrobarometer towns to the closest mission and to the closest mission with a printing press. . . . .	182
A.4	Summary statistics: Afrobarometer data . . . . .	184
A.5	Historical and Geographical Characteristics of Printing Presses With and Without a Publication Record . . . . .	185
A.6	Investments of Printing Presses With and Without a Publication Record . .	186
A.7	Estimation of the propensity score (logit) . . . . .	188
A.8	Propensity score matching: Balancing test (Printing Press) . . . . .	189
A.9	Propensity score matching: Balancing test (School) . . . . .	190
A.10	Propensity score matching: Balancing test (Health) . . . . .	191
A.11	Distance to missionary societies and Contemporary outcomes, OLS estimation (100km restriction) . . . . .	194
A.12	Distance to mission investments and Radio and Television, OLS estimation (100km restriction) . . . . .	195
A.13	Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction), Reporting all the covariates . . . . .	196
A.14	Distance to mission investments and Contemporary outcomes, OLS estimation (100km restriction), Standardized coefficients (Beta Coefficients) . . . . .	197

A.15 Distance to printing presses and Contemporary outcomes, OLS estimation, No restriction on distance . . . . .	199
A.16 Distance to printing presses and Contemporary Outcomes, OLS estimation using a binary variable (100km restriction) . . . . .	200
A.17 Printing Press and Contemporary Outcomes, Probit Estimation . . . . .	202
A.18 Printing Press and Contemporary Outcomes, with Potential “Bad Controls” . . . . .	202
A.19 Distance to the Largest Missions and Contemporary Outcomes . . . . .	203
A.20 Printing Press and Contemporary Outcomes, Clustering at the Mission Level . . . . .	203
A.21 Printing Press and Newspaper Readership, Former British Colonies . . . . .	204
A.22 Proximity to a printing press and Contemporary outcomes:Additional Robustness with one-way clustering . . . . .	205
A.23 Proximity to a printing press and Contemporary outcomes: Additional Robustness with two-way clustering . . . . .	206
B.1 $F_{ST}$ and Other Measures of Cultural Distance in the ACS dataset . . . . .	210
B.2 Effect of Religious Distance and Ancestry-Inherited Regional Characteristics on Income . . . . .	213
B.3 Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - General Perceptions of Life Measure . . . . .	214
B.4 Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Work-Ethic Measure . . . . .	215
B.5 Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Family Values Measure . . . . .	216
B.6 Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - Religion and Morale Measure . . . . .	217
B.7 Effect of WVS-Based Cultural Distance and Ancestry-Inherited Regional Characteristics on Income - All Categories Measure . . . . .	218
B.8 Effect of $F_{ST}$ on income - Controlling for Average Polarization . . . . .	220



# Bibliography

- ACEMOGLU, D. (2005): “Constitutions, Politics, and Economics: A Review Essay on Persson and Tabellini’s the Economic Effects of Constitutions,” *Journal of Economic Literature*, 43, pp. 1025–1048.
- (2011): “Lecture Notes,” MIT Open Course Ware.
- ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): “The Colonial Origins of Comparative Development: An Empirical Investigation,” *The American Economic Review*, 91, pp. 1369–1401.
- ACEMOGLU, D., T. REEDS, AND J. ROBINSON (2014): “Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone,” *Journal of Political Economy*, 122, 319–368.
- ACEMOGLU, D. AND J. ROBINSON (2012): *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Profile Books.
- ACS (2010): “American Community Survey and Puerto Rico Community Survey 2010 Subject Definitions,” Tech. rep., Census Bureau.
- AGER, P. AND M. BRUKNER (2013): “Cultural diversity and economic growth: Evidence from the US during the age of mass migration,” *European Economic Review*, 64, 76–97.
- ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): “Public Goods and Ethnic Divisions,” *The Quarterly Journal of Economics*, 114, 1243–1284.
- ALESINA, A., D. DEVLEESCHAUWER, W. EASTERLY, S. KURLAT, AND R. WACZIARG (2003): “Fractionalization,” *Journal of Economic Growth*, Vol. 8, 155–194.
- ALESINA, A., W. EASTERLY, AND J. MATUSZESKI (2011): “Artificial States,” *Journal of the European Economic Association*, 9, 246–277.
- ALESINA, A. AND E. LA FERRARA (2000): “Participation in Heterogeneous Communities,” *The Quarterly Journal of Economics*, 115, 847–904.
- (2005): “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, 43, 762–800.

- ALGAN, Y. AND P. CAHUC (2010): “Inherited Trust and Growth,” *The American Economic Review*, 100, pp. 2060–2092.
- (2014): *Handbook of Economic Growth*, Handbooks of Economics Series, chap. Trust, Growth, and Well-Being: New Evidence and Policy Implications, 49–120.
- ALGAN, Y., C. HEMET, AND D. LAITIN (2016): “The Social Effects of Ethnic Diversity at the Local Level: a Natural Experiment with Exogenous Residential Allocation,” *Journal of Political Economy*, 124.
- ALSAN, M. (2015): “The Effect of the TseTse Fly on African Development,” *American Economic Review*, 105, 382–410.
- 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, 151–184.
- ANDERSON, B. (1991): *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, Verso Books.
- ANGRIST, J. D. AND J. S. PISCHKE (2009): *Mostly harmless econometrics: an empiricist’s companion*, Princeton University Press.
- ASHRAF, Q. AND O. GALOR (2013): “The ‘Out of Africa’ Hypothesis, Human Genetic Diversity, and Comparative Economic Development,” *American Economic Review*, 103, 1–46.
- BAI, Y. AND J. K.-S. KUNG (2011): “Diffusing Useful Knowledge While Spreading God’s Message: Protestantism and Economic Prosperity in China, 1841-1920,” *Mimeo, The Hong-Kong University of Science and Technology*.
- BAIRD, S. J., R. S. GARFEIN, C. T. MCINTOSH, AND B. ÖZLER (2012): “Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial,” *The Lancet*, 379, 1320–1329.
- BANERJEE, A. AND E. DUFLO (2011): *Poor Economics*, Penguin Books.
- BARBUJANI, G. AND R. R. SOKAL (1990): “Zones of sharp genetic change in Europe are also linguistic boundaries,” *Proceedings of the National Academy of Sciences*, 87, 1816–1819.
- BAUER, T., G. S. EPSTEIN, AND I. N. GANG (2005): “Enclaves, Language, and the Location Choice of Migrants,” *Journal of Population Economics*, 18, 649–662.
- (2006): “The Influence of Stocks and Flows on Migrants’ Location Choices,” *CEDI Working Paper*, 0613.

- BEACH, H., ed. (1903): *A Geography and Atlas of Protestant missions: Their Environment, Forces, distribution, methods, Problems, results and Prospects at the opening of the Twentieth century.*, Student Volunteer movement for foreign missions.
- BEACH, H. P. AND C. H. FAHS, eds. (1925): *World Missionary Atlas*, Institute of Social and Religious Research.
- BECKER, S. O. AND L. WOESSMANN (2009): “Was Weber Wrong? A Human Capital Theory of Protestant Economic History,” *Quarterly Journal of Economics*, 124, 531–596.
- BÉTHUNE, L. (1889): *Les Missions Catholiques d’Afrique*, Desclée.
- BHATT, S., D. J. WEISS, E. CAMERON, D. BISANZIO, B. MAPPIN, U. DALRYMPLE, K. E. BATTLE, C. L. MOYES, A. HENRY, P. A. ECKHOFF, E. A. WENGER, O. BRIET, M. A. PENNY, T. A. SMITH, A. BENNETT, J. YUKICH, T. P. EISELE, J. T. GRIFFIN, C. A. FERGUS, M. LYNCH, F. LINDGREN, J. M. COHEN, C. L. J. MURRAY, D. L. SMITH, S. I. HAY, R. E. CIBULSKIS, AND P. W. GETHING (2015): “The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015,” *Nature*, 526, 207–211.
- BISIN, A. AND T. VERDIER (2001): “The Economics of Cultural Transmission and the Dynamics of Preferences,” *Journal of Economic Theory*, 97, 298 – 319.
- BLANCHFLOWER, D. G. (2004): “Self Employment, More may not be Better,” Tech. Rep. 10286, NBER.
- BORJAS, G. J. (1986): “The Self-employment Experience of Immigrants,” *Journal of Human Resources*, 485–506.
- BRATTON, M., R. MATTES, AND E. GYMAH-BOADI (2005): *Public Opinion, Democracy, and Market Reform in Africa*, Cambridge University Press.
- BRITANNICA, E., ed. (2000): *Britannica Encyclopedia*, Encyclopaedia Britannica.
- BROWN, S., P. E. SAVAGE, A. M.-S. KO, M. STONEKING, Y.-C. KO, J.-H. LOO, AND J. A. TREJAUT (2013): “Correlations in the population structure of music, genes and language,” *Proceedings of the Royal Society of London B: Biological Sciences*, 281.
- CAGÉ, J. (2014): “Media Competition and the Provision of Information,” Mimeo, Harvard.
- CAGE, J. (2015): *The Oxford Handbook of Africa and Economics*, Oxford University Press, chap. The Economics of the African Media, 605–625.
- CAGE, J. AND V. RUEDA (2016): “The Long-Term Effects of the Printing Press in Sub-Saharan Africa,” *American Economic Journal: Applied Economics*, 8, 1–31.



- CANTONI, D. (2013): “The Economic Effects of the Protestant Reformation: Testing the Weber Hypothesis in the German Lands,” Discussion Papers in Economics 14811, University of Munich, Department of Economics.
- (2015): “THE ECONOMIC EFFECTS OF THE PROTESTANT REFORMATION: TESTING THE WEBER HYPOTHESIS IN THE GERMAN LANDS,” *Journal of the European Economic Association*, 13, 561–598.
- CARRINGTON, W. J., E. DETRAGIACHE, AND T. VISHWANATH (1996): “Migration with Endogenous Moving Costs,” *The American Economic Review*, 86, 909–930.
- CASEY, K. (2015): “Crossing Party Lines: The Effects of Information on Redistributive Politics,” *American Economic Review*, 105, 2410–48.
- CAVALLI-SFORZA, L., P. MENOZZI, AND A. PIAZZA (1994): *The History and Geography of Human Genes*, Princeton University Press.
- CAVALLI-SFORZA, L. L., A. PIAZZA, P. MENOZZI, AND J. MOUNTAIN (1988): “Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data,” *Proceedings of the National Academy of Sciences*, 85, 6002–6006.
- CHANDLER, T. (1987): *Four Thousand Years of Urban Growth: An Historical Census*, St. David’s University Press.
- CHURCH MISSIONARY SOCIETY, ed. (1896): *The Church Missionary Atlas*, Student Volunteer movement for foreign missions.
- (CIA), C. I. A., ed. (2000): *The World Factbook 2000*, Stationery Office.
- CLAIR, C. (1976): *A History of European Printing*, Academic Press.
- CLARK, W. (2003): *Immigrants and the American Dream: Remaking the Middle Class*, Guilford Press.
- COGNEAU, D. AND A. MORADI (2014): “Borders That Divide: Education and Religion in Ghana and Togo Since Colonial Times,” *The Journal of Economic History*, 74, 694–729.
- COHEN, J. AND P. DUPAS (2010): “Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment,” *The Quarterly Journal of Economics*, 125, 1–45.
- COHEN, J., P. DUPAS, AND S. SCHANER (2015): “Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial,” *American Economic Review*, 105, 609–45.
- COLLIER, P. AND A. HOEFFLER (2004): “Greed and grievance in civil war,” *Oxford Economic Papers*, 56, 563–595.

- COMAS, D., F. CALAFELL, E. MATEU, A. PÉREZ-LEZAUN, E. BOSCH, R. MARTÍNEZ-ARIAS, J. CLARIMON, F. FACCHINI, G. FIORI, D. LUISELLI, D. PETTENER, AND J. BERTRANPETIT (1998): “Trading Genes along the Silk Road: mtDNA Sequences and the Origin of Central Asian Populations,” *The American Journal of Human Genetics*, 63, 1824 – 1838.
- COMPANY, C. (1911): *The Century Atlas: Africa [map]*, Buffalo, NY: Matthews-Northryp.
- CROZET, M. (2004): “Do migrants follow market potentials? An estimation of a new economic geography model,” *Journal of Economic Geography*, 4, 439–458.
- CUTLER, D., W. FUNG, M. KREMER, M. SINGHAL, AND T. VOGL (2010): “Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India,” *American Economic Journal: Applied Economics*, 2, 72–94.
- D’ALPOIM GUEDES, J., T. BESTOR, D. CARRASCO, R. FLAD, E. FOSSE, M. HERZFELD, C. C. LAMBERG-KARLOVSKY, C. M. LEWIS, M. LIEBMANN, R. MEADOW, N. PATTERSON, M. PRICE, M. REICHES, S. RICHARDSON, H. SHATTUCK-HEIDORN, J. UR, G. URTON, AND C. WARINNER (2013): “Is Poverty in Our Genes? A Critique of Ashraf and Galor, “The ‘Out of Africa’ Hypothesis, Human Genetic Diversity, and Comparative Economic Development,” *American Economic Review* (Forthcoming),” *Current Anthropology*, 54, 71–79.
- D’ALPOIM GUEDES, J., D. REICH, M. HERZFELD, N. PATTERSON, T. BESTOR, D. LIEBERMAN, J. COMAROFF, R. FLAD, D. CARRASCO, H. SHATTUCK-HEIDORN, K. LAMBERG-KARLOVSKY, J. COMAROFF, R. MEADOW, M. LIEBMANN, G. URTON, D. PILBEAM, A. KLEINMAN, AND B. GOOD (2012): “Response to Ashraf and Galor ‘The Out of Africa Hypothesis, Human Genetic Diversity and Comparative Economic Development’,” *SSRN*, 1–2.
- DAUBERT, P. (2009): *La presse écrite d’Afrique francophone en question: essai nourri par l’essor de la presse française*, Etudes Africaines, Paris: l’Harmattan.
- DE GRUCHY, J. W. (1999): *The London Missionary Society in Southern Africa: Historical Essays in Celebration of the Bicentenary of the LMS in Southern Africa, 1799-1999*, David Philip Publishers.
- DELL, M. (2010): “The Persistent Effects of Peru’s Mining Mita,” *Econometrica*, 78, 1863–1903.
- DELLA VIGNA, S. AND E. LA FERRARA (2015): *Handbook on the Economics of the Media*, Elsevier, chap. Economic and Social Impacts of the Media, 723–766.
- DENNIS, J. S., H. P. BEACH, AND C. H. FAHS, eds. (1903): *World Statistics of Christian Missions*, Student Volunteer movement for foreign missions.

- DEPETRIS-CHAUVIN, E. AND D. N. WEIL (2013): “Malaria and Early African Development: Evidence from the Sickle Cell Trait,” Working Paper 19603, National Bureau of Economic Research.
- DESMET, K., I. ORTUÑO-ORTÍN, AND R. WACZIARG (2015): “Culture, Ethnicity and Diversity,” Tech. Rep. 20989, NBER.
- DIAMOND, J. (2010): *Guns, Germs and Steel*, W.W. Norton.
- DITTMAR, J. E. (2011): “Information Technology and Economic Change: The Impact of The Printing Press,” *The Quarterly Journal of Economics*, 126, 1133–1172.
- DOUGLAS, M. (1986): *How Institutions Think*, Frank W. Abrams lectures, Syracuse University Press.
- DUFLO, E., P. DUPAS, AND M. KREMER (2015): “Education, HIV, and Early Fertility: Experimental Evidence from Kenya,” *American Economic Review*, 105, 2257–2297.
- DUPAS, P. (2011): “Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya,” *American Economic Journal: Applied Economics*, 3, pp. 1–34.
- EISENSTEIN, E. L. (1980): *The Printing Press as an Agent of Change*, no. vol. 1 in Volumes 1 and 2 in One, Cambridge University Press.
- ELLIS, W. (1844): *History of the London Missionary Society*, John Snow.
- ENGERMAN, S. L. AND K. L. SOKOLOFF (1997): *How Latin America Fell Behind*, Stanford University Press, chap. Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States.
- ESTEBAN, J., L. MAYORAL, AND D. RAY (2012): “Ethnicity and Conflict: An Empirical Study,” *American Economic Review*, 102, 1310–42.
- ESTEBAN, J. AND D. RAY (1994): “On the measurement of polarization,” *Econometrica*, 62, 819–819.
- (1999): “Conflict and distribution,” *Journal of Economic Theory*, 87, 379–415.
- FAHS, C. H. AND H. E. DAVIS (1935): *Conspectus of Cooperative Missionary Enterprise*, International Missionary Council.
- FARIA, N. R., A. RAMBAUT, M. A. SUCHARD, G. BAELE, T. BEDFORD, M. J. WARD, A. J. TATEM, J. D. SOUSA, N. ARINAMINPATHY, J. PÉPIN, D. POSADA, M. PEETERS, O. G. PYBUS, AND P. LEMEY (2014): “The early spread and epidemic ignition of HIV-1 in human populations,” *Science*, 346, 56–61.

- FEYRER, J. AND B. SACERDOTE (2009): “Colonialism and Modern Income: Islands as Natural Experiments,” *The Review of Economics and Statistics*, 91, 245–262.
- FISCHER, G., H. VAN VELTHUIZEN, AND F. O. NACHTERGAELE (2000): *Global Agro-Ecological Zones Assessment: Methodology and Results*, vol. IR-00-064, Laxenburg, Austria: IIASA - International Institute for Applied Systems Analysis.
- FRASER, R. (2008): *Book History through Postcolonial Eyes: Rewriting the Script*, Routledge.
- FULFORD, S. L., I. PETKOV, AND F. SCHIANTARELLI (2015): “Does it Matter Where You Came From? Ancestry Composition and Economic Performance of U.S. Counties, 1850-2010,” Discussion Paper 9060, Institut zur Zukunft der Arbeit (IZA).
- GALLEGO, F. A. AND R. WOODBERRY (2010): “Christian Missionaries and Education in Former African Colonies: How Competition Mattered,” *Journal of African Economies*, 3, 294–329.
- GENTZKOW, M. (2006): “Television and Voter Turnout,” *Quarterly Journal of Economics*, 121, 931–972.
- GENTZKOW, M., J. M. SHAPIRO, AND M. SINKINSON (2011): “The Effect of Newspaper Entry and Exit on Electoral Politics,” *American Economic Review*, 101, 2980–3018.
- GIULIANO, P., A. SPILIMBERGO, AND G. TONON (2014): “Genetic distance, transportation costs, and trade -super-1,” *Journal of Economic Geography*, 14, 179–198.
- GLAESER, E., R. LA PORTA, F. LOPEZ-DE SILANES, AND A. SHLEIFER (2004): “Do Institutions Cause Growth?” *Journal of Economic Growth*, 9, pp. 271–303.
- GLAESER, E. L. AND A. SHLEIFER (2002): “Legal Origins,” *The Quarterly Journal of Economics*, 117, 1193–1229.
- GORDON-BROWN, A. (1979): *The settlers’ press: seventy years of printing in Grahamstown covering the publication of books, pamphlets, directories, almanacs, newspapers, with historical notes and anecdotes and contemporary illustrations*, A. A. Balkema.
- GROSJEAN, P. (2014): “A History of Violence: The Culture of Honor and Homicide in the US South,” *Journal of the European Economic Association*, 12, 1285–1316.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2004): “The Role of Social Capital in Financial Development,” *American Economic Review*, 94, 526–556.
- (2006): “Does Culture Affect Economic Outcomes?” *Journal of Economic Perspectives*, 20, 23–48.
- (2008): “Long Term Persistence,” Working Paper 14278, NBER.

- HABERMAS, J. (1989): *The Structural Transformation of the Public Sphere: An Inquiry into a Bourgeois Society*, The MIT Press.
- HALPERIN, D. T. AND H. EPSTEIN (2007): “Why is HIV Prevalence so severe in Southern Africa?” *The Southern African Journal of HIV medicine*.
- HARDIMAN, D., ed. (2006): *Healing Bodies, Saving Souls*, The Wellcome Series in the History of Medicine.
- HARMSSEN, R. (1982): “The Printing Industry in sub-Saharan Africa. An Exploratory Study,” Ph.D. thesis, Rochester Institute of Technology.
- HUILLERY, E. (2009): “History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa,” *American Economic Journal: Applied Economics*, 1, 176–215.
- (2011): “The Impact of European Settlement within French West Africa: Did Pre-colonial Prosperous Areas Fall Behind?” *Journal of African Economies*, 20, 263–311.
- IDLER, E. L. (2014): *Religion as a social determinant of public health*, Oxford University Press.
- JOHNSON, H. B. (1967): “The Location of Christian Missions in Africa.” *Geographical Review*, 57, 168–202.
- KISZEWSKI, A., A. MELLINGER, A. SPIELMAN, P. MALANEY, S. E. SACHS, AND J. SACHS (2004): “A global index representing the stability of malaria transmission,” *American Journal of Tropical Medicine and Hygiene*, 70, 486–498.
- LAVAL, G., E. PATIN, L. B. BARREIRO, AND L. QUINTANA-MURCI (2010): “Formulating a Historical and Demographic Model of Recent Human Evolution Based on Resequencing Data from Noncoding Regions,” *PLoS ONE*, 5, e10284.
- LÉVI-STRAUSS, C. (1969): *The Elementary Structures of Kinship*, Beacon Paperback. no. BP 340, Beacon Press.
- LEVINSON, D. (1998): *Ethnic Groups Worldwide: A Ready Reference Handbook*, EBSCO eBook Collection, Oryx Press.
- LIVINGSTONE, D. (1858): *Missionary Travels and Researches in South Africa*, Harper and Bros.
- LUCAS, R. (2014): *International Handbook on Migration and Economic Development*., Elgar original reference, Edward Elgar Publishing Limited.
- MAAKE, N. (2000): *Politics of Publishing in South Africa*, University of Natal Press, chap. Publishing and Perishing: People, Books and Reading in African Languages, 127–162.

- MAJA-PEARCE, A. (1996): *Directory of African Media*, International Federation of Journalists.
- MCCLEARY, R. AND J. PESINA (2012): “Protestantism and Human Capital in Guatemala,” Mimeo, Harvard University.
- MICHALOPOULOS, S. AND E. PAPAIOANNOU (2011): “The Long-Run Effects of the Scramble for Africa,” Working Paper 17620, National Bureau of Economic Research.
- (2013): “Pre-colonial Ethnic Institutions and Contemporary African Development,” *Econometrica*, 81, 113–152.
- (2014): “National Institutions and Subnational Development in Africa,” *The Quarterly Journal of Economics*, 129, 151–213.
- MIGUEL, E. AND M. K. GUGERTY (2005): “Ethnic diversity, social sanctions, and public goods in Kenya,” *Journal of Public Economics*, 89, 2325 – 2368.
- MIGUEL, E. AND M. KREMER (2004): “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 72, 159–217.
- MONTALVO, J. G. AND M. REYNAL-QUEROL (2005): “Ethnic polarization, potential conflict, and civil wars,” *American Economic Review*, 796–816.
- MORAN, J. (1973): *Printing Presses: History and Development from the Fifteenth Century to Modern Times*, University of California Press.
- MUNSHI, K. (2003): “Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market,” *The Quarterly Journal of Economics*, 118, 549–599.
- MURDOCK, G. P. (1967): *Ethnographic Atlas: a Summary*, Pittsburg University Press.
- MYTTON, G. (1983): *Mass Communication in Africa*, Edward Arnold.
- NUNN, N. (2008): “The Long Term Effects of Africa’s Slave Trade,” *Quarterly Journal of Economics*, 123, 139–176.
- (2009): “Christians in Colonial Africa,” Mimeo, Harvard University.
- (2010): “Religious Conversion in Colonial Africa,” *American Economic Review Papers and Proceedings*, 100, 147–152.
- (2014): *Africa’s Development in Historical Perspective*, Cambridge University Press, chap. Gender and Missionary Influence in Colonial Africa.
- NUNN, N. AND L. WANTCHEKON (2011): “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 101, 3221–3252.

- OMU, F. I. A. (1978): *Press and Politics in Nigeria, 1880-1937*, Ibadan History Series, Longman.
- OSTER, E. (2005): “Sexually Transmitted Infections, Sexual Behavior, and the HIV/AIDS Epidemic,” *The Quarterly Journal of Economics*, 120, pp. 467–515.
- (2010): “Estimating HIV Prevalence and Incidence in Africa from Mortality Data,” *The B.E. Journal of Economic Analysis & Policy*, 10.
- (2012): “Routes of Infection: Exports and HIV Incidence in Sub-Saharan Africa,” *Journal of the European Economic Association*, 10, pp. 1025–1058.
- (2013): “Unobservable Selection and Coefficient Stability: Theory and Validation,” Tech. Rep. w19054, NBER Working Paper.
- PARKER, P. (1997): *National cultures of the world: a statistical reference*, ABC-Clio ebook, GREENWOOD Publishing Group Incorporated.
- PEMBERTON, T. J., M. DEGIORGIO, AND N. A. ROSENBERG (2013): “Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation,” *G3 (Bethesda)*, 3, 891–907, 23550135[pmid].
- PIKETTY, T. AND G. ZUCMAN (2014): “Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010,” *Quarterly Journal of Economics*, 129.
- PORTA, R. L., F. LOPEZ-DE SILANES, AND A. SHLEIFER (2008): “The Economic Consequences of Legal Origins,” *Journal of Economic Literature*, 46, 285–332.
- PUTNAM, R. D. (2000): *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon & Schuster.
- (2007): “E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture,” *Scandinavian Political Studies*, 30, 137–174.
- REDMAN, C. AND G. GROUP (1993): *Willings Press Guide 1993*, Willing’s Press Guide, Gale/Cengage Learning.
- (2003): *Willings Press Guide 2003*, Willing’s Press Guide, Gale / Cengage Learning.
- (2012): *Willings Press Guide: 2012*, Cision UK.
- RELETFORD, J. H. (2012): *Foundation of Human Biology : Human Population Genetics*, Wiley-Blackwell.
- ROSS, R. M., S. J. GREENHILL, AND Q. D. ATKINSON (2013): “Population structure and cultural geography of a folktale in Europe,” *Proceedings of the Royal Society of London B: Biological Sciences*, 280.

- ROWLING, C. F. AND C. WILSON (1923): *Bibliography of Christian Literature*, London Conference of Missionary Societies of Great Britain and Ireland.
- RUBIN, J. (2014): "Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation," *The Review of Economics and Statistics*, 96, 270–286.
- SEN, A. (1999): *Development as Freedom*, Oxford University Press.
- SETEL, P., M. LEWIS, AND M. LYONS (1999): *Histories of Sexually Transmitted Diseases and HIV/AIDS in Sub-Saharan Africa*, Contributions in medical studies, Greenwood Press.
- SNYDER, J. AND D. STRÖMBERG (2010): "Press Coverage and Political Accountability," *Journal of Political Economy*, 118, 355–408.
- SOCIO-ECONOMIC DATA AND APPLICATION CENTER (SEDAC) (2016): "Global Roads Open Access Data Set," .
- SPOLAORE, E. AND R. WACZIARG (2009): "The Diffusion of Development," *The Quarterly Journal of Economics*, 124, 469–529.
- (2015): "Ancestry, Language, and Culture," Working Paper 2142, NBER.
- STANLEY, B. (2009): *The World Missionary Conference, Edinburgh 1910*, Studies in the History of Christian Missions, Eerdmans Publishing Company.
- STOCK, E. (1899): *History of the Church Missionary Society*, Church Missionary Society.
- STRÖMBERG, D. (2004): "Radio's Impact on Public Spending," *The Quarterly Journal of Economics*, 119, 189–221.
- SWITZER, L. (1984): "The African Christian Community and its Press in Victorian South Africa." *Cahiers d'études africaines*, 24, 455–476.
- SWITZER, L. AND D. SWITZER (1979): *The Black Press in South Africa and Lesotho: a Descriptive Bibliographic Guide to African, Coloured, and Indian Newspapers, Newsletters, and Magazines, 1836-1976*, Bibliographies and guides in African studies, Hall.
- TABELLINI, G. (2010): "CULTURE AND INSTITUTIONS: ECONOMIC DEVELOPMENT IN THE REGIONS OF EUROPE," *Journal of the European Economic Association*, 8, 677–716.
- THORNTON, R. L. (2008): "The Demand for, and Impact of, Learning HIV Status," *The American Economic Review*, 98, pp. 1829–1863.
- TISHKOFF, S. A., F. A. REED, F. R. FRIEDLAENDER, C. EHRET, A. RANCIARO, A. FROMENT, J. B. HIRBO, A. A. AWOMOYI, J.-M. BODO, O. DOUMBO, M. IBRAHIM, A. T. JUMA, M. J. KOTZE, G. LEMA, J. H. MOORE, H. MORTENSEN, T. B.



- NYAMBO, S. A. OMAR, K. POWELL, G. S. PRETORIUS, M. W. SMITH, M. A. THERA, C. WAMBEBE, J. L. WEBER, AND S. M. WILLIAMS (2009): “The Genetic Structure and History of Africans and African Americans,” *Science*, 324, 1035–1044.
- TUDESQ, A.-J. (1995): *Feuilles d’Afrique: étude de la presse de l’Afrique sub-saharienne*, Publications de la Maison des sciences de l’homme d’Aquitaine, Editions de la maison des sciences de l’homme d’Aquitaine.
- VALENCIA-CAICEDO, F. (2014): “The Mission: Human Capital Transmission, Economic Persistence and Culture in South America,” JMP, Pompeu-Fabra.
- VAN BERGEIJK, P. AND S. BRAKMAN (2010): *The Gravity Model in International Trade: Advances and Applications*, Cambridge University Press.
- VAUGHAN, M. (1991): *Curing their ills: Colonial power and African illness*, Polity Press.
- VOIGTLÄNDER, N. AND H.-J. VOTH (2012): “Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany,” *The Quarterly Journal of Economics*, 127, 1339–1392.
- WANTCHEKON, L., M. KLASNJA, AND N. NOVTA (2012): “Education and Human Capital Externalities: Evidence from Colonial Benin,” Mimeo, Princeton University.
- WEBER, M. (1930): *The Protestant Ethic and the Spirit of Capitalism*, London and Boston: Unwin Hyman.
- WEIL, D. N. (2010): “Endemic Diseases and African Economic Growth: Challenges and Policy Responses,” *Journal of African Economies*.
- WOODBERRY, R. D. (2004): “The Shadow of Empire: Christian Missions, Colonial Policy, and Democracy in Postcolonial Societies,” Ph.D. thesis, University of North Carolina at Chapel Hill.
- (2012): “The Missionary Roots of Liberal Democracy,” *American Political Science Review*, 106, 244–274.
- WOODBERRY, R. D. AND T. SHAH (2004): “Christianity and democracy: The pioneering Protestants,” *Journal of Democracy*, 15, 47–60.
- ZELL, H. M. (1980): *The African book world and press : a directory*, Zell London, 2nd ed.
- ZIMMERMAN, K. AND A. CONSTANT (2004): *How Labor Migrants Fare*, Springer, chap. Labor market assimilation and the self-employment decision of immigrant entrepreneurs, 191–222.