Infrastructures, Economic Development and Health

by

Mansoureh Abbasi

A thesis submitted in partial fulfillment of the requirements for the Doctorate in Philosophy Degree in Economics

Department of Economics
Faculty of Social Sciences
University of Ottawa

Abstract

,

Chapter 1- We offer the first causal evidence on the effect of high-speed Internet on sexual behaviour, and HIV spread in Africa. Exploiting the gradual connection of the Internet network to the submarine Internet cables in Africa's coastal regions in combination with individual-level data, we apply difference-in-differences and find that exposure to fast Internet reduces the likelihood of HIV infection in locations close to the Internet network. The effect is significantly larger for women, educated individuals, older adults, and urban residents. The analysis of causal mechanisms reveals that, while access to fast Internet increases the likelihood of early sexual initiation and extramarital relationships, it also increases knowledge of HIV prevention and transmission, the use of modern methods of contraception, reduces fertility, and the likelihood of being in a polygamous marriage. We do not find that employment status, occupational choice, or migration drives the effect of the fast Internet on HIV. In addition, we use google trends data to show that Internet access increases the search popularity of sex-related topics, contraception, and HIV/AIDS. The findings imply that the expansion of Internet networks in Africa requires educational policies to promote healthy and stable intimate connections.

Chapter 2- We combine longitudinal data on the financial activities of individuals with information on temporal and spatial variations in broadband diffusion in the United States in recent years to study the causal impact of high-speed Internet on financial decisions. Our identification strategy controls for individual fixed effects and uses lightning strike density as an instrument for broadband penetration. We find that broadband diffusion increases the probability of investment in the stock market, having a mortgage, and having an active savings account. Analyzing causal mechanisms, we find that access to broadband Internet increases financial literacy and the use of financial services such as online banking and having a financial advisor. Additionally, heterogeneity analysis shows that the effect of broadband penetration on stock market participation significantly differs by gender, age, marital status, education, income, level of social connection, and mental health status. Analyzing the interconnection between financial decisions provides novel insights into the well-known stock market participation puzzle in the United States. We show that the probability of investing in the stock market is lower for an individual with a mortgage payment. Also, the probability of participation in the stock market is higher for a person with active savings.

Chapter 3- Evidence for road expansion and electrification as drivers of job creation is limited and mixed, with most studies considering either one or the other and only in isolation. We estimate the average and heterogeneous impacts of road and electricity investments and the interaction of the two on job creation over the past two decades in 27 countries of sub-Saharan Africa. Exploiting the exogenous location of ancestral ethnic homelands, we introduce a new instrumental variable for road accessibility inspired by post-independence leaders' agenda of building roads to extend authority over their country's entire expanse and promote nation-building. we used the topography and lightning strikes, a key source of damage to electric lines and disruption of service, to instrument the electricity supply. We find that constructing new roads and electricity grids positively and significantly affects employment from higher access to roads and electric grids. Moreover, the interaction between access to the two enhances the effects, making them complementary investments. The impacts of both individual and bundled investments are positive, but with differences between men and women, workers of various ages, and countries at different stages of development. In urban areas, better access to roads and electricity promotes employment, mostly related to the skilled base job creation. In rural areas, greater access induces a transition from low-to-high-skilled occupations. These differential effects suggest that the structural transformation brought about by road and electricity expansion is primarily a rural phenomenon.

Declaration of Authorship

I, Mansoureh Abbasi, declare that this thesis titled "Infrastructures, Economic Development and Health" and its work are my own. I acknowledge the contribution of Roland Pongou to the research associated with all three chapters of this thesis. I also acknowledge the contribution of Bluette Arcady Mongoue, Mathilde Lebrand and Fan Zhang to the research associated with the third chapter. In all cases, their contributions are equal.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ________________________________

Date: 2023 February 30

General Introduction

Access to healthcare, education, and infrastructure is at the heart of efforts to achieve sustainable development. Given the importance of these factors, researchers and policymakers have given much attention to finding the proper combination of instruments and strategies to promote economic development. My thesis studies the economic and health impacts of different kinds of infrastructure. The first chapter examines the health effects of technological improvements in communications, focusing specifically on access to high-speed Internet in coastal African countries. The second chapter examines the extent to which households' financial knowledge and financial behaviour are affected by access to high-speed Internet in the United States. Finally, the third chapter examines how the labor market responds to expanding roads and electrical infrastructure in African countries.

Connection to submarine cables has provided more affordable and higher-speed Internet access to individuals in coastal African countries. In the first chapter of my thesis, we investigate how access to high-speed Internet through submarine Internet cables affects sexual behaviour and the spread of HIV in Africa. This analysis combines geographic data from the Demographic and Health Surveys (DHS) and AIDS Indicator Surveys (AIS) with geo-referenced data on fibre-optic cables from 2009 to 2012. Owing to the endogenous nature of Internet diffusion, and following the work of Hjort and Poulsen (2019), we interact the information on individuals' location connectivity to the Internet network with the timing of the arrival of submarine cables and use the difference-in-differences approach to establish a causal relationship between Internet access and HIV infection. We find that access to submarine cables decreased the probability of HIV infection in locations close to the Internet network. Interestingly, the effect is primarily driven by women, educated individuals above 24 years old, and urban residents.

We show that, after the arrival of high-speed Internet infrastructure, the use of the Internet on a daily or monthly basis increased among individuals living close to the Internet network. We also find that these individuals better understand HIV transmission modes and prevention methods. Further, we find that women living in these locations are more likely to use modern contraception methods, such as condoms, and have lower fertility rates than women living in unconnected locations. Importantly, women living in connected locations are less likely to be involved in a polygamous marriage. However, access to the Internet facilitates finding sexual partners, increasing the probability of having extramarital relationships and reducing the age of sexual debut among women. This chapter's findings imply that expanding Internet access in Africa requires educational policy interventions to optimize its use and promote healthy and stable intimate relationships.

In the second chapter of the thesis, we investigate the extent to which household financial literacy and financial decisions in the U.S. are affected by access to high-speed broadband Internet following the Federal Communication Commission's (FCC) 2010 decision to provide affordable access to robust broadband service with at least 4 Mbps of download speed and 1 Mbps of upload speed. To answer these questions, we merge individual-level data from the Health and Retirement Study (HRS) with data from the national broadband map of residential high-speed broadband diffusion in the United States. The HRS has information on the location (fips code), financial activities, demographics, and health statuses of US adults above the age of 45. The combined dataset includes data for the 2010 to 2016 period. Our baseline estimates are produced using OLS regressions with the county, year, cohort and individual fixed effects. To address the endogenous nature of Internet diffusion, we use lightning strike density per square kilometre as an instrument for broadband diffusion. Unlike previous studies in the literature, our work examines the effects of Internet diffusion from the supply side. We are also the first in the literature to control for unobserved individual heterogeneity using individual fixed effects. The findings indicate that Broadband diffusion increased Internet use and financial literacy among US adults. Specifically, we find that an increase in broadband diffusion is associated with an increase in the number of individuals who participate in the stock market, invest in housing, and have active checking or savings accounts. We also find that, for individuals with an outstanding mortgage, the effect of broadband diffusion on the probability of investing in the stock market and having active checking or savings accounts is more muted. Moreover, for individuals with active checking or savings accounts, the effect of broadband diffusion on the probability of buying stocks and investing in housing is higher than for other individuals. Furthermore, people advise more with financial specialists and use online banking services via the Internet. Further results show the probability of having stocks and mutual funds decreases for older people, retired individuals, individuals with lower educational levels, individuals who are not married, depressed, and have lower levels of income. This chapter's findings shed light on the well-known stock market participation puzzle in the United States.

In the third chapter, we analyze the complementary role of road and electricity infrastructure in job creation and structural change in sub-Saharan Africa. Most of the previous works study the effect of each infrastructure on job creation separately. Very few studies have investigated the interaction between different types of infrastructure and their joint effects on the labour market. In this chapter, we investigate the joint effects of road and electricity infrastructure on job creation in 27 sub-Saharan African countries. We also examine how social and demographic factors mediate these effects. We use a least-cost path algorithm, assuming that land altitude is the primary geographic determinant of construction costs for roads and construct hypothetical roads linking ancestral ethnic homelands to the capital of each country. Our novel approach considers proximity to these hypothetical roads as an instrument for road access. We also instrument the proximity to an electricity grid with lightning strike density and land elevation, both of which are exogenous variables. Using an OLS approach with time and country fixed effects and our instrumental variable approach, we compare the probability of finding a job for individuals living close to road and electricity infrastructure with those not close to either infrastructure.

Findings indicate that investment in road networks and electricity grids has positive and significant effects on job creation. Combined with road networks and electricity access, their interaction strengthens the labor supply. We also find that the benefits of investing in these forms of infrastructure, either simultaneously or separately, in terms of job creation are greater in more developed African economies. Within countries, we find that the benefits of expanding these forms of infrastructure vary by location. Specifically, roads contribute more to job creation in rural areas, while electricity has a greater impact in urban areas. We find that employment gains due to infrastructure investment accrue to women to a greater degree than men and to younger adults more than older adults. Additionally, investment in these forms of infrastructure changes the structure of the labour market, increasing employment in skilled occupations and decreasing it in unskilled occupations (primarily self-employed jobs in agriculture). This chapter suggests that a significant factor in sub-Saharan Africa's development could be the joint development of road and electricity infrastructure, as the effect of one boosts the effect of the other. Therefore, it could be advantageous to combine infrastructural development with poverty reduction policies that target improvements in the labour market in this region.

Acknowledgments

I am grateful to be on this journey that has transformed me into a different person. I would like to express my sincere gratitude to God for guiding me through this experience.

I am incredibly thankful to my supervisor, Professor Roland Pongou. His invaluable advice, unwavering support, and patience have been instrumental throughout my Ph.D. journey. His extensive knowledge and wealth of research experience inspired me from the very beginning. It has been an honor to work with and learn from him. I cannot adequately express my heartfelt appreciation for his time, ideas, and contributions to my work.

I would also like to extend my sincere thanks to my thesis committee members: Professor Myra Mohen, Professor Jose Galdo, and Professor Myra Yazbeck. Their insightful comments and valuable assistance have greatly improved my research and enriched my professional experience. I am also grateful to the external examiner, Professor Eugene Beaulieu, for dedicating his time and providing valuable feedback on my thesis.

I am deeply grateful to the faculty and staff of the University of Ottawa's Economics Ph.D. program, particularly Dr. Louis-Philippe Morin, the program director, and Dr. Catherine Deri-Armstrong, the head of the Department of Economics, for their unwavering support. I would also like to thank Dr. Louis Hotte for his kind assistance and prompt responsiveness. Additionally, I would like to express my appreciation to the administrative staff, including Diane Ritchot, Carole-Anne Gratton, Irene Pare, Guilherme Queiroz, and Martine Raymond, for their consistent support and assistance.

I want to acknowledge and express my gratitude to the friends who have been with me throughout this journey. Their companionship, discussions, and shared learning have enriched my experience. Special thanks to Jin Wang, Arcady Mongoue, Landry Kuate Fotue, Fatima Zohora, Lamis Kattan, Marie Christelle Mabeu, and my dear sisters Farzaneh Davarzani, Modjgan Alishahi, and Azadeh Farahnakian.

Lastly, I owe this accomplishment to my parents and my sister, Mahboubeh. Their unconditional love, unwavering support, and constant encouragement have been the reasons behind all of my achievements.

To all ***Iranian women and girls*** *who courageously demonstrate for their fundamental human rights.*

Contents

List of Tables

List of Figures

Chapter 1

AIDS in the Digital Age: Causal Evidence from Submarine Internet in Africa

1.1 Introduction

The digital revolution has produced a structural change in societies, impacting multiple sectors ranging from economic production to commerce and culture. The Internet has redefined and bypassed most traditional media due to lower communication costs. While the considerable societal benefits of the digital revolution in terms of the accessibility of information are widely acknowledged, its implications are still not well understood, especially in the context of developing countries. In Africa, Internet use and access are still low, but the Internet market is one of the fastest-growing in the world. Over the last decade, the number of Internet nodes has doubled in this region, and the share of individuals with Internet access has grown from 2 percent in 2000 to over 46 percent in 2020. Despite this fast-growing adoption rate, research into its causal impacts on African economies and societies has just begun with the work of Hjort and Poulsen (2019). We still know little about how the Internet has affected family and household behaviors. This knowledge gap constitutes a significant constraint on the design of policies intended to optimize the use of information and communication technologies in developing societies.

This paper provides evidence of the causal impact of access to *high-speed* Internet on sexual behavior and HIV/AIDS in Africa. To understand the possible mechanisms underlying this effect, we consider how Internet access affects knowledge of HIV transmission modes and prevention methods, premarital and extramarital sexual behavior, sexual debut age, marriage, fertility, and polygamy. These outcomes are interesting in their rights, as their analysis provides significant insights into how the digital revolution has affected family outcomes. In addition, we document significant heterogeneity in the effects of Internet access on these outcomes. We primarily consider heterogeneity by gender, given significant gender differences in sexual behavior, HIV/AIDS, and their related consequences in Africa (Pongou and Serrano (2013)). We also document heterogeneous effects by age, education, and place of residence. The analysis of the heterogeneous effects of fast Internet sheds additional light on channels through which this technology affects sexual behavior, and HIV spread in African countries.

Theoretically, Internet access has an ambiguous effect on HIV spread.[1] As a readily available and low-cost communication technology, the Internet provides access to a wide range of

[1]Throughout the paper, while the theoretical argument speaks to the potential impacts of the Internet on sexual behavior and HIV spread in general, our empirical analysis mainly investigates the effect of *submarine* Internet. The arrival of submarine cables in Africa's coastal regions significantly increased Internet speed and decreased the cost of access to online information.

Internet-based sexual education programs available at almost any time and place for various age groups. Additionally, the Internet makes sharing experiences, providing social support, and promoting protective contraception methods easier for individuals. Anonymity on the Internet also allows individuals to find answers to their health questions, discuss any sexual issues, and collect knowledge about healthy sexual behaviors. Hence, it can be used as an effective tool to mitigate the spread of HIV. At the same time, by reducing search costs, the Internet has also become an increasingly common and effective way of finding a sexual partner, which increases sexual connections between people from different backgrounds (Rosenfeld and Thomas (2012); Schwartz (2013)). Moreover, exposure to Internet pornography has potential implications for an individual's sexual behavior (Bhuller et al. (2013)) and attitudes toward using contraception and having multiple sexual partners (Koletić et al. (2019)). Consumption of online pornography can therefore raise the likelihood of an individual participating in risky sexual behaviors, exacerbating HIV spread.

Africa offers an ideal setting for our analysis. The HIV epidemic, one of the world's most significant public health issues, is primarily concentrated in this region. Of the estimated 38.4 million individuals living with HIV in 2020 (including 1.8 million children), about 25.6 million lived in Africa (UNAIDS (2020)). The broad toleration of men's sexual activity in African societies, along with women's lower bargaining power and limited knowledge of HIV transmission modes and prevention methods, has made the possibility of HIV contraction more pronounced among women (Pongou and Serrano (2013)). In 2020, women and girls accounted for approximately 59% of all new HIV infections (UNAIDS (2020)). The vast majority of HIV infections in Africa are generated through having unprotected heterosexual contact, having multiple sexual partners, living in polygamous families, and starting sexual relations at an early age. The enduring HIV/AIDS epidemic has subtle effects on African economies; these effects are mainly produced through lowering labour productivity, reducing savings and factor accumulation, and increasing health expenditures (Azomahou et al. (2016)). Numerous programs and policies have evolved to mitigate the spread of HIV in Africa. They are mainly school-based or other micro-level educational programs for specific groups or regions (Dupas (2011)) and frequently focus on providing education about HIV/AIDS prevention, transmission and misconceptions (Duflo et al. (2015a); Dupas (2011)). The Internet in Africa is accessible to a large segment of males and females, allowing us to assess how its impacts differ along important demographic and socioeconomic personal characteristics.

For our analysis, we combine the Demographic and Health Surveys (DHS) and the AIDS Indicator Surveys (AIS) with geo-referenced data on fiber-optic cables from 2009 to 2012 and information on the timing of submarine Internet cables' arrival in Africa's coastal regions. DHS and AIS are individual-level surveys that are nationally and sub-nationally representative. They collect data on a wide range of socioeconomic, demographic, and health outcomes, including information on sexual behavior, marital outcomes, and HIV status. Our main dataset for HIV infection consists of 79,587 women and 32,912 men.

The main challenge associated with identifying the causal effect of Internet access on sexual behavior and HIV/AIDS stems from the fact that the placement of Internet infrastructures may not be random. In African countries, these infrastructures might have been built in places with more substantial expected economic benefits for telecommunication companies or followed long-standing patterns in adopting other communication technologies. These endogeneity issues pose a fundamental problem in determining the causal effect of the Internet. In Africa's coastal regions, connecting Internet backbone networks to submarine Internet cables from Europe in the late 2000s created an ideal setting to overcome such endogeneity issues. Following a similar research design as in Hjort and Poulsen (2019), we estimate the causal effect of high-speed Internet access by exploiting the gradual arrival of these submarine cables and conducting a difference-in-differences analysis.

In African countries, Internet traffic in each country travels through its national backbone infrastructures constructed by national telecommunication companies. Starting in 2009, an additional connection of submarine cables to Internet backbone networks (Figure 1-4) improved speed and take-up rate (see Figure 1-7). Notably, countries were connected to the submarine cables separately and at different times, making it possible to assess the causal impact of the Internet in a difference-in-differences framework. Our treatment variable is generated by interacting a binary indicator for the presence of submarine cables in a country with a variable measuring an individual's geo-location connectivity to the fiber optic internet network.[2]

We find that access to fast Internet significantly reduces the likelihood of HIV infection. Individuals living close to the Internet backbone are over 2 percentage points less likely to become infected with HIV after the arrival of submarine cables. This effect is heterogeneous, mainly driven by effects on women, individuals over 24 years old, educated individuals, and urban residents. In particular, exposure to the fast Internet reduces the likelihood of HIV infection by 2.7 percentage points among women versus only 0.7 percentage points among men, with the latter effect statistically insignificant at the 10 percent level. The fact that the effect is more significant among educated individuals and urban residents shows that the health benefits of fast Internet in Africa are occurring more to individuals who are already

[2]Our analysis is robust to several alternative cutoffs, and by using a continuous connectivity measure. As in Hjort and Poulsen (2019), we restrict our primary analysis to observations within 10 km of the Internet backbone network to ensure that the treatment and control groups are comparable regarding characteristics likely to affect sexual behavior and HIV spread.

privileged, leading to an even greater digital divide on the continent. These heterogeneous effects indeed suggest that access to the fast Internet affects HIV status by providing reliable information on prevention methods, primarily to individuals who can afford access to such information given their education and their socioeconomic status.

We test several direct mechanisms through which connection to the submarine Internet might affect HIV status. First, we construct an index that measures comprehensive knowledge of HIV transmission modes and prevention methods. This index is generated based on answers to five questions asked to interviewees in the DHS and the AIS. The value of the index varies from 0 to 5, with a higher value indicating higher knowledge. Exposure to submarine Internet significantly raises HIV knowledge; almost all of this effect comes from women. Second, we find that connection to submarine Internet increases the use of modern contraception. The effect on condom use is larger for women (2 percentage points) than for men (1 percentage point), consistent with Internet access empowering the former more. Fast Internet also decreases fertility among women, which reflects greater use of contraceptive methods.

Third, we investigate the effect of connection to the submarine Internet on the likelihood of being involved in a polygamous marriage. We find that access to this technology decreases the probability of a woman marrying a polygamous man by over 3 percentage points; in the context of sub-Saharan Africa, where polygamy is still high, this is an economically significant effect. It has been shown that polygamous networks bias HIV infection against women (Pongou and Serrano (2013)). Reducing polygamy is an important channel through which fast Internet access decreases HIV infection among women.

While the channels documented above imply that fast Internet decreases HIV infection, we also find that access to this technology significantly increases sexual activities. This latter effect is larger for women. Among women, connection to the submarine Internet increases the possibility of having sexual partners other than the spouse by 2 percentage points. It decreases age at first sex by around half a year–a large effect. In addition, it lowers the probability of marriage among women and men, but these effects are not statistically significant. These findings suggest that fast Internet decreases the cost of finding male partners. The lower effects for men can be explained by the fact that men's sexual misconduct is tolerated in most African societies (Pongou and Serrano (2013)). So the Internet does not significantly increase their existing power.

We complement our study by collecting Google Trends data and combining it with the novel on the growth of Internet nodes over the last decade in African countries. Using this dataset, we show that Internet access increases the search popularity of sex-related topics (having sex, pornography, sexual words, sexual abuse), sexual partners, contraceptive methods, HIV/AIDS, and other sexually transmitted diseases. Findings confirm our analysis based on individual-level data, which shows that Internet access has two opposing effects on HIV prevention: an adverse effect through increasing sexual activities and a positive effect through increasing modern contraception and decreasing the likelihood of polygamous unions. Our analysis shows that the positive effect of submarine Internet dominates its negative effect; submarine Internet access in Africa significantly decreases the likelihood of HIV infection, especially for women.

Furthermore, given the documented effect of the fast Internet on employment in Africa (Hjort and Poulsen (2019)), it is important to determine the latter drives the effects of this technology on sexual behavior and HIV. We show that this is not the case. In fact, we show that these effects are robust to controlling for employment status and occupational choice. Also, we find that Internet access decreases HIV prevalence similarly among individuals who have a job and those who do not. Moreover, among individuals who have a job, Internet access decreases HIV prevalence both among individuals who have low-skilled jobs and those who have low-skilled jobs, with the effect being larger in the latter category. To check if migration does not drive the effect of fast Internet on HIV, I restricted the analysis to individuals who did not move ten years before the survey. Results represent a stronger effect for individuals who did not change their place of residence for 10 years before the survey. These findings demonstrate that the mechanisms that underlie the effect of the fast Internet on job creation and employment in Africa are quite distinct from the channels through the effect of this technology on sexual behavior and HIV operations.

We conduct several robustness checks to confirm the validity of our analysis. First, our findings are robust to changes in the radius around the Internet network that we use to define individual locations' connectivity, confirming that the effect of submarine Internet is not sensitive to how connectivity to the Internet backbone is defined. Second, we check if the development of other infrastructures, such as roads and electrical grids, paralleled the expansion of the backbone network and might therefore be driving the effect of fast Internet access on HIV. To this end, we use the maps of Africa's road and electricity networks to define a location's connectivity to each of these infrastructures. Controlling for these variables does not impact the effect of the fast Internet on HIV. Third, we use an alternative data set on the evolution of the Internet nodes from 2009 to 2019 to define Internet connectivity and find that our results are robust to this new measure of Internet accessibility. Finally, following Conley and Taber (2011), we compute the wild cluster bootstrap P-value and the 95% confidence interval to check if limited variation in the timing of arrival of submarine cables is a source of bias for our results. Allowing for arbitrary correlation of error terms within the same country, year, or individuals do not affect our findings.

The paper's findings suggest that expanding high-speed Internet infrastructures may be among the most feasible policy options to control HIV in Africa and may also improve community-based programs' efficacy. However, the adverse impact of this technology on some of the most important family outcomes suggests that its expansion should be accompanied by other educational policies that optimize its utilization.

The rest of this paper is organized as follows. Section 1.2 presents our main contributions to the related literature. Section 1.3 presents our conceptual framework, describing the theoretical channels through which Internet access might affect HIV. Section 1.4 describes the data used for the analysis, also providing background information on Internet development in Africa. Section 1.5 presents our empirical strategy. Section 1.6 presents the primary results. Section 1.7 discusses the mechanisms. Section 1.8 presents some robustness checks and sensitivity analyses for the identification strategy. Finally, Section 1.9 concludes the paper.

1.2 Contribution to the literature

This paper offers novel findings on the causal impact of access to high-speed Internet on sexual behavior and HIV spread in Africa. It also documents significant heterogeneity in the effect of this technology and identifies several mechanisms that may contribute to this underlying relationship. The different outcome variables, involved in the analysis of the mechanism, are interesting in their own rights and reveal potential benefits and drawbacks of fast Internet access in Africa.

Our paper contributes to the growing literature on the impacts of the Internet. Several studies have tested the effect of broadband Internet on different economic, political, and social outcomes, including buying and selling, employment, trade, investment, political activity, marriage, and criminal behavior. For example, broadband Internet has significantly affected trade in China by improving firm performance and facilitating communication between buyers and input suppliers (Fernandes et al. (2019)). In France, the Internet increased firm-level imports (Malgouyres et al. (2019)). The Internet has also been shown to affect voting behavior in Germany (Falck et al. (2014)) and participation in political protest in Russia (Enikolopov et al. (2020)). Internet access has also been found to increase racial hate crimes (Chan et al. (2016)). More related to our paper is Bhuller et al. (2013) document a positive effect of Internet access on sex crime, likely due to increased exposure to pornography.

The causal impact of the Internet has not been widely studied in Africa. Hjort and Poulsen (2019) find that submarine Internet increases employment in this region. They argue that high-speed Internet creates vast mobility in the job market, contributes to job creation, and enables skilled workers to find better-matched jobs with high payment. Lebrand et al. (2023) document significant complementary between high-speed Internet and the expansion of roads in creating jobs in sub-Saharan African countries. We differ from these studies in our research questions, conceptual framework, main findings, and mechanisms of impact. We investigate the effect of high-speed Internet on HIV and document opposing effects on variables likely to affect this outcome. We also find that the effect of high-speed Internet on HIV does not vary much by employment status or by type of employment, which implies that the operating channels of impact documented in the aforementioned studies are not the same as in our setting. Despite these differences, given the demonstrated positive impact of family planning on female labor participation (Miller (2010)), our finding that the fast Internet positively affects the use of modern methods of contraception and reduces fertility can be seen as providing another channel through which this technology positively affects employment outcomes as documented in Hjort and Poulsen (2019) and Lebrand et al. (2023).

We also contribute to the large literature on the determinants of sexual behavior, reproductive outcomes, and HIV spread in Africa. Given the high cost of antiretroviral therapies, HIV prevention is the best strategy to avoid the financial liability induced by HIV treatment for most African countries (Collier and Sterck (2018)). This makes it important to study the socioeconomic determinants of sexual behavior and HIV spread. Canning et al. (2020a) find that colonial origins and population policies affect sexual behavior, marriage, and fertility outcomes in sub-Saharan Africa. The risk of HIV infection has been found to be affected by fertility behavior (Yao (2022)), education and wealth (Fortson (2008); Case and Paxson (2013); De Walque (2009); Alsan and Cutler (2013); De Walque (2007); Iorio and Santaeulàlia-Llopis (2016)), spousal search behavior (Magruder (2011)), premarital sexual relations (Meekers and Calves (1997)), transactional sex (Epstein and Morris (2011)), and sibling sex composition (Mongoue and Pongou (2023)). Other documented determinants of HIV spread include the slave trade (Bertocchi and Dimico (2019)), legal origins (Anderson (2018)), gender norms and sexual network structure (Pongou and Serrano (2013)), marriage (Greenwood et al. (2017)), ethnic heterogeneity (Pongou (2009)), and information about HIV transmission (Dupas (2011); Duflo et al. (2015a)), among others.[3] The literature has also focused on how HIV education programs are likely to empower women. Meta-analyses generally report weak impacts of such programs, regardless of whether they are presented in

[3]These studies imply that Africa differs from other settings like the United States in terms of the determinants of sexual behavior (Adamopoulou (2013).

the classroom (McCoy et al. (2010); Duflo et al. (2015b)) or as peer-provided information (Cornish and Campbell (2009)). Other studies find that the impacts of education programs are significant (Arcand and Wouabe (2010); Dupas (2011)). Our paper shows that access to high-speed Internet increases knowledge about HIV transmission modes and prevention methods, increases the use of modern contraception, decreases fertility, and decreases involvement in polygamous relationships among young women. Although high-speed Internet increases sexual activities, our analysis suggests its positive effect on HIV knowledge and prevention dominates, significantly decreasing HIV infection.

1.3 Conceptual framework

We argue that the theoretical effect of high-speed Internet access on sexual behavior and HIV spread is ambiguous. Our conceptual framework, summarized by Figure 1-5, offers two opposing channels through which Internet access may affect HIV. By making information about health and HIV transmission modes and prevention methods quickly accessible at low cost, high-speed Internet can encourage the use of protective contraception and foster healthy sexual behavior, therefore reducing exposure to HIV/AIDS. Before the Internet's introduction, sexual health information was mainly available through magazine articles, television, library books, and healthcare facilities, making access to information expensive and time-consuming. The Internet, by contrast, does not only lower the cost of access to sexual information (Dorado et al. (2019)), but it also facilitates anonymous discussions about sex-related issues. This is especially significant given cultural contexts where such discussions with family members, school teachers, and friends are difficult or even shameful (Moran (2009)). Indeed, the ability to search for online information anonymously allows individuals, who would otherwise be reluctant to rely on family or friend networks, to obtain sexual health information. Various online social networks and chat rooms can introduce environments where individuals can exchange health-related messages. Facebook is the most popular social network, followed by MySpace and Twitter. Studies reveal that the most common search topics by youth on the Internet and social networks are HIV/STIs, pregnancy/childbirth, sexual acts/behavior, contraception, relationships/social issues, sexual abuse, services/research information, sexual pleasure, sexual terminology, reproductive cancers, and teen sexuality/virginity (Barman-Adhikari and Rice (2011); Dorado et al. (2019)). These studies, therefore, suggest that exposure to healthy messages via the Internet as a helpful information source can help control HIV prevalence.

Along the same lines, high-speed Internet can increase the possibility of finding better matches among other Internet users in several ways. First, it can reduce information constraints. For example, dating sites allow individuals to search for potential partners based on personal characteristics such as age, education, and profession, and less readily observed characteristics such as sexual preferences, hobbies, and religion. Second, the Internet may also expand an individual's choice set, making matches available that may have been previously unavailable due to a lack of connection in traditional social networks (Stevenson and Wolfers (2007)). Finally, the anonymity that the Internet offers can also expand the pool of people looking for a match. All of these circumstances can help in having safer sexual relations and controlling HIV spread.

While the Internet can help prevent HIV, its use may also increase risky sexual behavior and exposure to HIV in several ways. The proportion of young people exposed to pornographic materials through the Internet is growing (Igbo et al. (2013)). Access to the Internet exposes adolescents to sexually explicit content, leading to increased and risky sexual activity and to sexual violence (Bhuller et al. (2013); Garofalo et al. (2007)). By reducing the time cost of finding partners, high-speed Internet access facilitates sexual links with multiple individuals and reduces the age at first sexual intercourse. These risky behaviors can contribute to a higher probability of contracting HIV.

Studies conducted in African countries confirm the conflicting effects that Internet access can have on HIV spread. A study by Wusu (2013) reveals that individuals in Nigeria exposed to the Internet and sexual content for two hours are three times more likely to use condoms than those who spend one hour a day on different social media platforms. At the same time, spending two hours a day on the Internet doubles the probability of having more than one sexual partner. In Ethiopia, a study of preparatory school adolescents reported that more than 77% of the students knew about sexually explicit media (SEM) (Kimemia and Mugambi (2016)). Moreover, most of them (75.5%) had watched SEM movies, and more than 73% were exposed to sexually explicit texts. In Kenya, the Internet was found to increase access to sexually explicit music, images, videos, and texts. Similarly, access to social media networks increased online sex requests and contributed to high levels of teenage pregnancy among adolescent girls and young women in high schools (Kimemia and Mugambi (2016)).

It follows that high-speed Internet access has conflicting effects on HIV infection. While high-speed Internet access is likely to increase fast access to information on STIs and safe sexual practices, it may also encourage risky sexual behavior by lowering the cost of access to sexually explicit content and the cost of partner search. These ambiguous theoretical effects make it important to assess the causal effect of this technology on sexual behavior and HIV empirically. In addition to conducting this analysis, we document the possible mechanisms underlying this relationship. The next section documents the context and data used for the analysis.

1.4 Context and data

1.4.1 High-speed Internet in Africa

As information and communication technologies (ICTs) improve, broadband Internet and mobile phone technologies will continue to play an increasingly important role globally in social and economic processes. Introducing these technologies drives a range of benefits, including enhanced social connectivity and increased information transmission. A lack of fixed-line infrastructure in rural and landlocked areas has long hampered the development of fixed-line broadband Internet markets in Africa. Approximately 20 years ago, Internet access on the continent was minimal, with satellite Internet being the only available method of accessing the Internet. Having satellite Internet was costly, with users paying approximately ten times more for Internet use than in other regions. Africa still has the lowest number of Internet connections per capita globally, and network infrastructure is still in its early stages of development. At the same time, the Internet market in Africa is one of the fastest-growing in the world. According to the International Telecommunication Union, in 2000, Africa's Internet connectivity was less than that of Manhattan in New York City. In 2006, the continent had only 2% of Internet connections worldwide. By 2013, 13 percent of Africans had an Internet connection through a fixed line; in 2019, this number had increased to 22%. As of 2019, Africa's total Internet connectivity was 39.6 percent compared to 58.8 percent in the rest of the world. In 2021 the rate reached 46.20% compared to 64.70% for the world. Appendix table 1-A.6 compares the Internet growth and penetration rate in several African countries between 2000 and 2019. Among these countries, Nigeria had the highest penetration rate, with 96% of the population having access to the Internet.

Telecommunication companies in Africa are looking for broadband wireless access technologies as the key to making fast, reasonably priced Internet available to the population at large. Last-mile" Internet infrastructure in Africa consists of fibre cables, copper cables, and cell towers, with satellites connecting individuals to the Internet at large. In the period preceding the introduction of last-mile Internet, Internet traffic travelled through a national backbone infrastructure. National telecom companies in all African countries constructed the backbone infrastructure. The backbone network's main channels were constructed many decades ago, with the rest slowly expanding over time; in Figures 1-1 and 3-3, we compare the evolution of Internet nodes between 2009 and 2019, which displays 92 percent growth over this time period.

In 2007, the International finance corporation (IFC) and the World Bank launched an ambitious effort to build a network of submarine Internet cables (SMCs) off the coast of East Africa. These cables establish instant telecommunications link continental Internet networks and behave as information highways. They allow for international communication at high speeds and volumes. More than 320 fibre optic submarine cables transmitted 99% of all international telecommunications between 1990 and 2017. However, Africa was isolated for much of this period and had no submarine cable connections until 2009. Since then, the digital infrastructure has quickly been improved. Today, almost all coastal African countries are directly connected to the global Internet through SMCs. Typically one connection is made for each country passed by the cable (Hjort and Poulsen (2019)). The African SMCs are directly linked to southern Europe and Brazil.[4] Connections to the submarine Internet have lowered broadband costs considerably. Terrestrial backbones have been used as infrastructure for the submarine Internet.[5] With the submarine cables, faster speed and traffic come to Africa from other continents to locations close to the terrestrial network. The lack of connections between neighbouring countries' backbone networks means that each country has a unique submarine Internet treatment date, which we exploit using a difference-in-differences analysis to identify the causal effect of submarine Internet on sexual behavior and HIV. Figure 1-5 shows the evolution of these submarine lines between 2009 and 2012 in Africa's coastal areas. Connection to the submarine Internet has changed the socio-economic, cultural and behavioral structures of the societies on the continent. However, it has also increased the spatial digital divide between coastal and urban populations within the subcontinent. People who lived close to submarine cable landing stations and other critical backbone infrastructures gained access to a faster and more stable telecommunication network than their inland counterparts (Cariolle (2018)). This paper uses the arrival of eight submarine cables in Africa's coastal area between 2009 and 2012 as a treatment effect to study how submarine Internet access affects HIV infection and sexual behavior.

1.4.2 HIV in Africa

The HIV epidemic is an unprecedented global public health issue and a significant constraint on human capital accumulation. Africa has been the most severely affected continent. The first recorded case of HIV/AIDS in Africa was reported in 1983, and by 1986 it was clear that HIV had spread into the populations of many countries in Africa (Quinn et al. (1986)). According to the Joint United Nations Programme on HIV/AIDS (UNAIDS (2020)), of the estimated 38.4 million people infected with HIV/AIDS at the end of 2021, 68% were in Africa.

[4]The Brazilian cables eventually connect to the US east coast

[5]A single backbone network covers each country.

This region only constitutes about 10% of the world's population. The ten countries with the highest level of HIV prevalence as a percentage of the people are South Africa (25%), Nigeria (13%), Mozambique (6%), Uganda (6%), Tanzania (6%), Zambia (4%), Zimbabwe (6%), Kenya (6%), Malawi (4%) and Ethiopia (3%). The African Region also accounts for 70% of all new HIV infections. In 2021, about 1.1 million new people were infected with HIV (WHO, 2021). Nearly 80 percent of all HIV-positive women in the world live in Africa. Uniquely, it is the only place in the world where more women than men live with HIV (Pongou and Serrano (2013)). On average, adult women are three times more likely than men to be affected, and women between the ages of 15-24 are as much as eight times more likely than men to be HIV positive.

The high HIV prevalence on the continent has been primarily linked to the continent's specific socio-economic situation and cultural attitudes toward sexual relationships (Pongou and Serrano (2013); Anderson (2018)). The vast majority of HIV infections in Africa are generated through unprotected heterosexual contact (UNAIDS (2020)), having multiple sexual partners, living in polygamous families, and starting sexual relations at an early age. In South Africa and Kenya, HIV prevalence is almost three times higher in men who had sex with men than men who had sex with women only (Dunkle et al. (2013)). Sex work has also been the main reason for the epidemic in the region. The burden of HIV is very high among this group (Baral et al. (2012)). Injected drug use is another source of concern. The absence of intense reduction programs and persistent high-risk behaviors keep the risk of HIV transmission at a high level in the region. Every year, a substantial amount of national and international funding is directed toward eradicating HIV. Success in HIV prevention in Africa has the potential to mitigate the global burden of HIV. By documenting the causal effect of high-speed Internet on HIV, our study will likely inspire policy actions to slow the spread of HIV in this region.

1.4.3 Data and descriptive statistics

Individual-level data: We use Demographic and Health Surveys (DHS) and AIDS Indicator Surveys (AIS) of the DHS program for our analysis. These surveys are repeated cross-section surveys collected in most developing and middle-income countries since the mid-eighties. They are illustrative at the national and sub-national levels. Information on demographic and socioeconomic status (including age, education, and occupation) is collected for all household members. In addition, selected women and men provide information on a wide range of other variables, such as marriage, fertility, mortality, family planning, reproductive health, child health, nutrition, and HIV/AIDS. Due to the survey's subject matter, women of reproductive age (15–49) are the survey's focus. Hence the number of women in the selective samples is higher. The DHS Program routinely collects geo-referenced data (longitude and latitude) in selected surveyed countries, making it possible to combine these surveys with a wide range of external data and information. DHS uses a two-stage probabilistic sampling technique to select clusters (census enumeration units) in the first stage and households in the second stage. In recent surveys, respondents' blood samples were collected to test various health conditions, including HIV status. By collecting blood for HIV testing, the DHS Program provides nationally representative estimates of HIV prevalence rates. The testing is anonymous, voluntary, and non-informative to respondents.[6] Our sample for the HIV prevalence is restricted to DHS surveys containing HIV testing information and GPS coordinate data before and after the introduction of the submarine Internet. Using the information on each enumeration unit's geographical coordinates, we matched individual-level data to their locations relative to the terrestrial Internet network. These conditions restrict the sample to 112,499 individuals consisting of 79,587 women and 32,912 men across five countries: Congo (DRC), Ghana, Kenya, Mozambique and Tanzania.[7] To estimate the causal effects of the Internet, we only use the sample of individuals within 10 kilometers of the Internet backbone network. The restriction is similar to the one implemented in Hjort and Poulsen (2019); its purpose is to ensure that individuals treated through exposure to the submarine Internet are comparable to those not treated. This leaves our sample with 38,527 observations. 1-A.2 describes the socio-economic, HIV test results and sexual behaviour of people in the sample. At the time of the survey, ages ranged between 15 and 64 for women and between 15 and 59 for men. Women comprise around 75 percent of the sample because the DHS program over-samples women. The mean age for women and men is 29.28 and 30.15 years old, respectively. About 83 percent of women and 82 percent of men in the sample have a primary education level and 11 percent does not have any education. Individuals in the sample most frequently live in urban areas (56 percent), and 90 percent of them are exposed to any type of media other than the Internet, such as television, radio, and magazine. Around 73 percent of individuals had a job during the last 12 months preceding the survey. The plurality of individuals is Muslim (25 percent) or Christian (23 percent). In order to control for urbanization effects on HIV, we connect geo-referenced decadal population estimates of cities in 33 African countries from Africapolis with the closest enumeration unit at the cluster level in the DHS and control for the population density of the nearest city to each cluster. The mean population density is around 3,400 individuals per square kilometer for the whole sample. To check if the individual's decision to change the place of the resident

[6] The average response rate is very high; it is 93 percent for women and slightly lower for men.

[7] See Appendix table 1-A.1 for details.

affects our result, we define a dummy variable equal to one if the individual did not change the place during the last 10 years before the survey. This data is not available for Congo, Kenya, Ghana and Tanzania. In the sample, 41 percent of the respondents did not change their homes.

Table 1-A.2 also, summarizes information about HIV prevalence in selected countries. We define a dummy variable for HIV status, which takes the value of one if an individual's blood test result was HIV positive, and the value of zero otherwise. On average, 8.2 percent of individuals in the sample are infected with HIV, 9.2 percent of them are women and 4.9 percent are men. Among the countries in the sample, Mozambique has the highest infection rate (16 percent), followed by Kenya (9 percent). HIV prevalence is around 11 percent among women in urban areas and 5.5 percent for women living in rural areas. Also, HIV prevalence is over 12 percent for women older than 24 years.

To evaluate how high-speed Internet has affected HIV knowledge in Africa, we construct an index to measure this outcome. Using a similar approach as in Chan and Tsai (2018), we consider five questions on transmission modes and prevention methods incorporated in the Demographic and Health Surveys (DHS). These questions, which the United Nations Program considers as core indicators of HIV knowledge (UNAIDS (2020)), are the following:

- Can the risk of HIV transmission be reduced by having sex with only one faithful uninfected partner?
- Can the risk of HIV transmission be reduced by using condoms?
- Can a healthy-looking person have HIV?
- Can a person get HIV from mosquito bites?
- Can a person get HIV by sharing a meal with someone who is infected?

To construct our index of HIV knowledge, we give a respondent one point for a correct answer to a question and zero points otherwise. We then sum up the number of points across all the questions to obtain the index. Therefore, a respondent's score ranges from zero to five, depending on the number of correct answers to the questions.

In the sample, Ninety-eight percent of the respondents have heard about HIV, 72 percent knew that the probability of contracting HIV is reduced with condom use, and 80 percent knew that having only one sexual partner reduces the chance of HIV infection. Misconceptions about HIV are low, with 78 percent of the respondents believing that a healthy-looking person can have HIV; 77 percent believe that mosquitoes cannot transmit HIV, and 69 percent believe HIV cannot spread through sharing food. The HIV knowledge index has a value of 3.7 out of 5, with 3.98 for male and 3.7 for female respondents.

As we explained before, access to the fast Internet can affect different sexual behaviors. We defined different variables related to the sexual behaviours that might influence by access to the Internet and affect HIV contraction . First, we define a dummy variable for being in a polygamous marriage. For each sexually active woman, this variable is equal to one if she is married to a man who has other wives, and it is equal to zero otherwise; for each man, this variable takes the value of one if he has more than one wife, and zero otherwise. In our sample, around 16 percent of individuals are involved in a polygamous marriage, and the share of such individuals is more than two times greater for women (20 percent) than for men (8 percent). Another important sexual behaviour that could be affected by the arrival of the submarine Internet is the use of modern contraception during sexual acts. We define a dummy variable equal to one if an individual uses a condom during sexual relations, and equal to zero otherwise. At the time of the survey, around 35 percent of individuals in the sample were using contraceptives to avoid pregnancy; this rate was higher for men. Modern contraceptive use (including pills, IUD, injection(s), vaginal methods, condoms, female or male sterilization, and implants) is more prevalent than the use of traditional methods (including periodic abstinence (of any kind), withdrawal, lactational amenorrhea, and any other country-specific traditional method). The use of modern contraception within 10 kilometres of the Internet network is prevalent among 29 percent of women and 33 percent of men. Around 12 percent of individuals use a condom, with a higher rate among men (20 percent) than women (8 percent). We also consider extramarital relationships as another channel through which submarine Internet could affect HIV spread. We define a dummy variable equal to one if an individual is married or in a relationship with a main sexual partner and has another sexual partner, and equal to zero otherwise. Around 27 percent of individuals reported having sexual partners other than their spouses. This rate was higher for men (37 percent) than women (23 percent). Access to the fast Internet could also affect the age of sexual initiation. The average age of sexual initiation is close to 17.5 years old for women and 18.19 for men.

The fast Internet could also affect marriage and fertility patterns. We define a dummy variable equal to one if an individual is married at the survey, and zero otherwise. The marriage rate in the whole sample is around 60 percent; this rate was higher for women (63 percent) than men (51 percent). We consider fertility behavior, as a proxy for sexual behavior and the use of contraception. In order to define the fertility variable, we first merge the Children's record files of the DHS with the Individual (or Women) record files. We define a dummy variable equal to one if a woman was pregnant or gave birth to a child during the last year before the survey, and equal to zero otherwise. The share of women who were pregnant or had a birth during the last year before the survey was around 14 percent for the whole sample, and 10 percent for women within 10 kilometres of the Internet backbone network.

Distance to the Internet backbone and Internet nodes: To estimate an individual's geographic distance from their closest Internet network, we use African Terrestrial Fibre Optic Cable's (AfTerFibre 2014) map of terrestrial backbone networks (see Figure 1-3), This calculation is performed with QGIS. In the DHS and the AIS, the closest distance of an individual's location to the Internet backbone is calculated as the distance from the cluster's centroid in which that individual lives to the closest Internet network. To check the robustness of our main result, We also use data on long-haul optical fibre nodes between 2009 to 2019 from Africa Bandwidth Maps. The map displays Africa's terrestrial, satellite and submarine cable transmission networks. This map was provided to us by AfTerFibre. [8] Using the same approach, we compute the distance between each optical fiber node and the centroid of individual enumeration locations. Figures 1-1 and 1-2 show the evolution of Internet nodes from 2009 to 2019.

We consider an individual location connected to Internet when the distance between the centroid of enumeration location from the closest Internet network or optical fiber node is at most 1.5 kilometers. The idea behind the choice of 1.5 km for the location's connectivity comes from the fact that fiber optic cables are used for short-distance signal transmissions in concessions, campuses or apartment buildings. It also has the characteristic of guiding several modes simultaneously, allowing the routing of a greater volume of data simultaneously. It is generally used for Multi-mode fiber optic cables that can transmit a signal of 100 Mbit per second up to 2 kilometres. The single-mode optical fiber has a single mode of transmission. It has the characteristic of possessing very low signal attenuation, which optimizes the signal transmission speed. With this type of cable, the transmission distance can reach 5 kilometers. However, due to its high cost, it is mainly used for long-distance signal transmission. We define a dummy variable equal to one if an individual lives within 1.5km of each network and zero otherwise. In our sample, 39 percent of individuals live within 1.5 km of the internet network, and 8 percent live close to the fiber optic nodes.

Treatment variable:

To estimate the treatment effect of access to submarine Internet, we define the treatment and control groups using a location's distance from the 2009 Internet backbone network. We consider an individual's location as *connected* if they live within 1500 meters from the

[8] AfTerFibre is an open initiative to gather and share information about terrestrial fibre optic cable projects in Africa. Data on the location of Internet nodes from 2009 to 2019 are provided to us by this organization.

terrestrial backbone network and *unconnected* if they are farther than 1500 meters. This cutoff is used to have a sufficient number of observations. In robustness checks, we use an alternative radius and show that our results are not sensitive to these changes. Using the information on the arrival date of submarine cable for each country, we define an individual as *treated* if that individual lives in a location that is *connected* after the arrival of submarine Internet. In other words, in a country, a connected location becomes treated after the arrival of the submarine cable. A location that is not connected is in the control group both before and after the arrival of the submarine cable. This approach is natural and follows from Hjort and Poulsen (2019).

We generate 10km-by-10km grid cells and control for the grid-cell fixed effect to account for time-invariant heterogeneity at the local level; we also control for the interaction between each grid cell and an individual's location connectivity to the Internet. We compare our outcome variables for the treatment and control groups before and after the arrival of submarine cables in each country.

Internet take-up rate and Speed: Data on Internet speed comes from Hjort and Poulsen (2019). This study uses quarterly data released by Akamai on mean Internet connection speeds for 900 African locations during the 2007–2013 time period. Akamai averages the speeds recorded for residential users, educational institutions, government offices, and firms in each quarter's location. Our paper restricts analysis to location-quarters where Internet speed is measured using more than ten unique IP addresses. Before the arrival of submarine cables, the average Internet speed for connected regions was 452 kilobytes per second. After connection to submarine Internet cables, this was raised to 1800 Kilobytes per second. In unconnected locations, the Internet speed did not change significantly.[9].

In order to compare the change in Internet take-up rate before and after the arrival of submarine cables, we use locational data from Afrobarometer on the frequency of Internet use as a daily or any use of the Internet during a year. We define a dummy variable equal one if a respondent uses the Internet at least once a day. For any Internet use, the dummy variable is equal to one if an individual uses the Internet at least once a month. Before the arrival of submarine cables, around 7 percent of people in the data set used the Internet for everyday use. After the arrival of submarine cables, 11 percent of people used the Internet. Before the arrival of submarine cables, any use of the Internet during a month concerned 17 percent of the sample, with this figure increasing to 21 percent of the sample after the introduction of submarine Internet.

[9]Following Hjort and Poulsen (2019), we exclude the four most prominent cities in each country from the speed data sample.

Google Trends data: In order to investigate the effect of Internet expansion on search interest for sex-related topics, we used Google Trends data. Google Trends is a public data set that aggregates billions of search topics that someone types in the Google search box[10], and it has been considered as an indicator of people's awareness in different economic, environmental, and resources management topics (Choi and Varian (2012); Baker et al. (2011); Gonzales and Ajami (2017); Kam et al. (2019)). The data shows how popular a particular search term is over a specific time relative to the maximum volume of search activities given the chosen time. Trends present the search interest data at various global, national, state, metropolitan, and city levels. Trends are declared on a scale from 0 to 100, with 100 serving as the highest. A value of 50 is a topic searched half as often as the most popular term for the terms queried within the selected region and specific time frame, and zero is the lowest.

We aggregate monthly time-series data corresponding to the eleven selected African coastal countries between 2009 and 2019 to the yearly level. We consider twelve queries related to the possible mechanisms through which Internet access may affect HIV status to estimate how people's awareness regarding HIV and other sexual behaviors changes with Internet penetration. [11] Our panel data set consists of 121 observations for each word searched on Google.[12]

1.5 Methodology

1.5.1 ICT effects and endogeneity issue

In studying the effects of information and communication technologies (ICTs), one should acknowledge the fact that these technologies are generally not randomly allocated within a country. ICTs are likely to be introduced in locations with higher expected economic benefits and larger consumption markets. A profit-maximizing telecommunication firm prefers to roll out Internet infrastructures in areas where most individuals are willing to pay for an Internet subscription. Typically, these areas have a higher average income, a larger share of high-skilled workers, and a younger population. Given that these socioeconomic factors are usually correlated with sexual behavior, endogeneity can bias the estimated results (Bellou (2015); Agrawal (2021); Falck et al. (2014)).

[10] http://trends.google.com/trends

[11] These queries are HIV/AIDS infection, marriage, pregnancy test, sex, condom use, sexual abstinence, sexual abuse, sexual words, sexually transmitted diseases, polygamy, sexual partner, and porn movies.

[12] Note that Google Trends data are not available at the subnational level for most African countries. For this reason, we use yearly data at the country level, which complements our primary analysis that is based on individual-level data sets.

To solve this possible endogeneity problem, we exploit the fact that submarine cables originating from Europe to Africa only connect to the national backbone networks at the landing points at specific times. Since a single backbone network covers each country, there is no spillover effect in which one coastal country's submarine connection could affect neighbouring countries (Hjort and Poulsen (2019)). Hence, each country has a particular treatment date when the first submarine cable is plugged in, independent of the country's sexual behavior and HIV conditions. Quasi-random variation generated by the date of arrival of submarine Internet cables at landing points helps overcome the endogeneity problems stemming from the nature of ICT development, allowing to estimate of the causal effect of submarine Internet on sexual behavior, HIV, and other pertinent outcome variables.[13]

Our central analysis compares the HIV status of individuals within a specific distance from the Internet network with that of individuals living further away before and after the arrival of submarine Internet for five countries in Africa's coastal region. Following our conceptual framework, we consider two main groups of outcome variables that affect HIV spread and might be affected by Internet access; this will identify the mechanisms underlying the effect of submarine Internet on HIV status. These two groups of variables consist of general information about HIV and contraceptive use and sexual behaviors (including premarital and extramarital sexual interactions, marriage, and polygamy). This later analysis is conducted using individual information in five African coastal countries before and after the submarine cables' arrival.[14] A clear definition of the treatment and control groups was needed to find the treatment effect of submarine Internet on the outcome variables. This paper's definition of the treatment and control groups was derived using data on individuals' geographic distance from the Internet backbone network.

1.5.2 Identification strategy

The causal effect of submarine Internet is estimated with the following baseline difference-in-differences model:

$$h_{ij(i)c(i)t} = \alpha + \beta(T_{c(i)t} \times Connected_i) + \gamma_{j(i)} \times Connected_i + X'_{j(i)c(i)t}\theta + \delta_{c(i)t} + \epsilon_{j(i)c(i)t} \quad (1.1)$$

Where $h_{ij(i)c(i)t}$ is the value of an outcome variable for individual i, in grid-cell j, country c, and time t. The treatment variable, $T_{c(i)t}$, is a dummy variable equal to one if the Internet

[13] Hjort and Poulsen (2019) make similar assumptions when estimating the effect of submarine Internet on job creation and employment in Africa.

[14] These countries are represented in table 1-A.1 These are the only coastal sub-Saharan African countries for which DHS data on HIV were collected before and after the arrival of submarine cables.

backbone network in country c is connected to at least one submarine cable at time t. $Connected_i$ is a dummy variable equal to one if individual i resides within 1500 meters of the Internet backbone network. Our parameter of interest, β, is the causal effect of connection to the submarine Internet on the outcome variable. The interaction between ($\sim 10 \times 10km$) grid-cells, $\gamma_{j(i)}$, and $Connected_i$ controls for any time-invariant differences in the outcome variable that may be correlated with access to submarine Internet within each cell. The country-year fixed effect, $\delta_{c(i)t}$, controls for any within country and time variation differences in the outcome variable correlated with access to fast Internet. Standard errors are clustered at the grid-cell level. $X'_{i(i)jc(i)t}$ is the individual level vector of variables that control for any other non-treatment effects which may affect the outcome variables and are not correlated with treatment.

The central identification assumption is that prior to the submarine Internet connection, areas within 1500 meters and further than 1500 meters from the terrestrial backbone network had parallel trends in terms of the outcome variables, and did not experience any other systematically differential shocks after the arrival of submarine Internet. In section 1.8, possible issues with these identifying assumptions are examined. Results are not significantly affected by varying the radius around the backbone network used to define the connectivity status. They are also not significantly affected by varying the size of the grid cells used to define location-fixed effects.

Finally, to check if the treatment effect runs parallel to other infrastructures such as roads or electricity, we use maps of Africa's road and electricity network to define each location's connectivity to the road and electricity based on the distance to these infrastructures.[15] We interact the submarine cables treatment with proximity to roads and electricity networks. If locations near such infrastructures saw a faster reduction in HIV over time, irrespective of whether they were also connected to the Internet backbone, there is a risk of attributing the effect of these infrastructures to submarine Internet. For robustness checks, we also use an Internet node map to define connectivity and check if the findings are similar to those obtained for the effects of submarine Internet.

1.5.3 Baseline characteristics: Socio-economics, sexual behavior, and HIV

To verify the plausibility of the difference-in-differences method in our context, we compare the baseline characteristics of individuals living in locations close to the backbone network

[15] Abbasi et al. (2022) use similar data to study the complementarity role of roads and electricity in job creation in sub-Saharan African countries.

with those of individuals in locations farther away prior to the arrival of the submarine cables (Table 1-1). Panel A of Table 1-1 reports the results comparing baseline socioeconomic and demographic characteristics, including age, gender, education level, employment status, religion, and media exposure. We do not find economically or statistically significant differences in these characteristics. For example, the average age is 29.17 years in connected locations versus 29.59 years in non-connected locations. Educational attainment is a bit higher in connected locations, but the difference is small and insignificant. Around 73 percent of individuals in the sample had a job in the last 12 months preceding the survey, with no significant difference between the connected and the non-connected groups (76 vs. 73 percent). Around 29 percent of the individuals were Muslim (28 vs. 30 percent), and 27 percent were Christian (29 vs. 25 percent). Exposure to the media was a little bit higher in connected locations (95 vs. 92 percent), and the difference is significant.

Panel B of table 1-1 reports the results comparing baseline outcomes measuring Internet use, sexual behavior, HIV status, and other outcomes considered in the conceptual framework. Prior to the introduction of submarine Internet cables, roughly 23 percent of individuals were using the Internet; the difference in the take-up rate of the Internet between the connected and the non-connected locations (23 vs. 21 percent) was small and statistically insignificant. HIV prevalence was slightly higher in connected locations, 10 vs. 7 percent. Comprehensive knowledge of HIV prevention methods and transmission modes was high but did not differ between connected and non-connected locations (3.8 vs. 3.6). The use of modern protective contraception was, on average, low: only 25 percent of individuals used these methods, and it varied very little across connected and non-connected locations (28 vs. 26 percent). Condom use did not vary either (13 vs. 12 percent) nor did the use of traditional contraception (5.5 vs. 5.4 percent). Around 16 percent of women were married to a polygamous man, and this behavior was only marginally greater in connected locations (16 vs 14 percent). Of most sampled individuals, 71 percent had sexual partners (of any kind), and 24 percent had sexual partners other than their spouses. The probability of having a sexual partner of any kind was not significantly different in the connected and unconnected locations, but the probability of having a sexual partner other than one's spouse was 5 percentage points higher in connected locations. Age at first intercourse was around 18 years old and did not significantly differ between the two groups. The marriage rate in connected regions was 3 percent lower in locations close to the network. Finally, fertility choice during the last 12 months did not differ significantly in connected and not connected locations(11 percent vs 12 percent). In all, baseline characteristics differ very little between individuals in connected locations and those in non-connected locations, which justifies the use of the difference-in-differences method to identify the causal effect of submarine Internet on HIV and to study the underlying mechanisms of this relationship.

1.6 Findings

1.6.1 Effects of submarine Internet: Graphical illustrations

In this section, we show graphical illustrations of the effects of the submarine Internet. Figures 1-6, and 1-7 represent the effect of submarine Internet on HIV incidence, sexual behaviors, the Internet take-up rate and speed in coastal regions of Africa. Figure 1-6-a represents the average likelihood of getting infected with HIV declined by 10 percent (of the initial level) in connected areas following the arrival of submarine Internet, while it declined by only 1 percent in non-connected areas. Similarly, comprehensive knowledge of HIV prevention methods and transmission modes improved by 3.7 vs. 1.9 percent in connected vs. non-connected areas (Figure 1-6-b). Figure 1-6-c shows that condom use during sexual interactions increased by 15 percent following the introduction of submarine Internet in connected locations, while in non-connected areas, it only increased by 1.6 percent. As for polygamy, it decreased by 12.5 percent in connected locations versus only 5 percent in non-connected locations (Figure 1-6-d). The number of sexual partners other than the spouse increased by 10 percent in connected locations, while in not connected areas, it declined by a negligible 0.08 percent (Figure 1-6-e). Also, age at first sex declined 2 vs. 0.6 percent in connected and unconnected locations, respectively 1-6-f).

Change in Internet speed in areas within 1500 meters of the backbone network and areas located more than 1500 meters away from the backbone network before and after the arrival of submarine cables is shown in figure 1-7. While speed was comparable in both areas before the introduction of submarine Internet, it increased considerably in areas close to the backbone network afterwards. The Internet speed increased by 297 percent in connected areas versus only 89 percent in non-connected areas. Similarly, the Internet take-up rate improved by 40 percent in connected areas versus only 13 percent in non-connected areas (Figure 1-7-b).

1.6.2 HIV status

In this section, we present the findings about the effect of access to submarine Internet on the likelihood of being infected with HIV. We estimate this effect for the whole sample(table 1-2, then separately for women and men (tables 1-3 and 1-4. In each table, controls are added sequentially, which makes it possible to gauge the stability of the coefficient on the primary predictor. In Column (1), we only include the treatment variable and control for country-year and grid-cell fixed effects. Our estimates show that the probability of contracting HIV decreases by 2.2 percentage points for individuals in connected areas after the arrival of submarine cables; however, this effect is mainly driven by women. Indeed, access to submarine Internet reduces HIV infection by roughly 2.7 percentage points among women versus only 0.07 percentage points among men, with the latter effect being statistically insignificant.

In columns (2) to (6) of each table, other control variables are added. In Column (2), we control for the respondent's age in the survey. The effect of access to submarines increases slightly and remains statistically significant for the whole sample and women. We also note that age positively affects the likelihood of being infected with HIV. Being older by one year increases HIV infection by 0.2 percentage points, with this effect being slightly greater for women compared to men.

We control for employment status in column (3) of the tables. The effect of the treatment variable changes very little and remains much greater for women compared to men. The estimates indicate that the effect of access to submarine Internet on HIV is mediated neither by age nor by having access to a job. This is interesting, especially in light of the findings of previous studies, which show that submarine Internet positively affects job creation and employment in Africa (Hjort and Poulsen (2019)) and that poverty affects sexual behavior and HIV spread (Shannon et al. (2015)). This latter study argues that poverty increases vulnerability to HIV and sexually transmitted infections by driving women into sex work, where sex is exchanged for money, food, or other commodities. For men, the effect of poverty can be reversed, especially in a context in which women sell sex to men. Quite interestingly, we find that having access to a job decreases a woman's probability of contracting HIV, but this has no effect on men.

In column (4), we control for education, which does not seem to influence the effect of submarine Internet access on HIV infection. There is no consensus in the literature about the direction of this effect, and the evidence is mixed (Berhan and Berhan (2015)). Some early studies found higher rates of HIV among educated people, especially women, in several sub-Saharan African countries. Education affects women's sexual behavior by postponing marriage. Also, most educated women with better economic situations refuse to be involved in polygamous relationships and instead choose to have multiple partners, which might increase HIV risk (Dallabetta et al. (1993); Case and Paxson (2013)). However, recent studies find a strong and positive relationship between education and lower HIV contraction (Agüero and Bharadwaj (2014); De Neve et al. (2015)). Our results for the whole sample and the sample of women confirm that by increasing the education level from primary education to the secondary and high education level, the probability of HIV infection declines considerably.

In column (5) of the tables, we control for exposure to mass media such as television, radio, outdoor media (such as billboards and posters), and print media (such as magazines and newspapers). Theoretically, mass media exposure has an inconsistent effect on sexual behavior. Exposure to sexually explicit media promotes risky sexual behavior among youth (Wright (2011)). At the same time, media-based campaigns that attempt to promote various health behaviors to the public have had some success. Outside of HIV prevention, media-based campaigns have targeted tobacco use, heart disease, alcohol use, illicit drug use, cancer screening and prevention, sexual health issues, and other public health issues (Wakefield et al. (2010)). We find that media exposure has an overall negative effect on HIV infection. Similar to the effect of the submarine Internet, the effect of media exposure is driven only by women. Notably, the effect of media exposure among women is smaller than that of submarine Internet.

In column (6), we control for religious affiliation, which affects neither the size nor the statistical significance of the effect of access to submarine Internet on HIV. It has been argued that religious organizations can influence sexual behavior and the perceptions of HIV/AIDS in a community. Promoting health-related behaviors related to religious norms and participating in a religious network may help prevent the spread of HIV (Agadjanian (2001); Takyi (2003); Gray (2004)). As well, the church's emphasis on conservation and intense social closeness (e.g. youth groups, frequent prayer meetings) reduces the likelihood of Christians engaging in extramarital and premarital relationships (Anderson (2018)). If access to the Internet strengthens or weakens religious beliefs, it could negatively or positively affect HIV prevention and exposure. Our analysis shows very little evidence of this happening, as the coefficient on treatment does not change significantly after controlling for religion. Interestingly, religion itself has a negative effect on HIV infection, although this effect is not statistically significant.

In column (7), we control for the effect of wealth on HIV prevalence. There is a long-held belief that HIV epidemics are driven by poverty (Butler (2000); Fenton (2004)). It is believed that poverty results in the adoption of high-risk sexual behaviors, increasing the risk of acquiring HIV. Specific risky sexual practices that have been associated with poverty include earlier sexual debut (Hallman (2005)) and reliance on sex work in order to generate income. Other studies have challenged the idea that poverty fuels HIV epidemics and have shown that HIV prevalence can be higher in wealthier populations (Johnson and Way (2006)). More affluent individuals may engage in other risky sexual practices, such as more regular or casual sex partners, and those who are infected might survive longer due to greater access to treatment and care and healthier diet options (Braitstein et al. (2006)). DHS questionnaires ask about what the households own based on an extensive list of assets and other housing characteristics such as Productive assets, Non-productive assets, and Household utilities that reflect their economic status. To control the wealth effect, we construct an index of wealth by considering 39 questions in the DHS surveys reflecting the economic status of African society. Table 1-A.6 represents the summary statistics of the variables that are considered in the wealth index. We use principal component analysis (PCA) to make a comparable index of wealth among the countries. [16] Our measure of wealth is the first principal component. It explains most of the variance. Concerning the wealth effect. Our finding represents that increased wealth will reduce HIV prevalence among women and men. However, this effect is not significant.

In column (8), we control for the effect of urbanization on HIV prevalence. It has long been known that HIV infection is higher at the urban than rural levels. The urbanization process, the rise in the proportion of a country's total population that lives in urban areas, directly connects with HIV/AIDS prevalence. The positive association between HIV and population density. This statement is probably true for many countries. Explanations for this positive relationship include the fact that more populated regions serve as a conduit for new influences through the growth in injecting drugs, higher rates of social interaction, late marriage, and, more importantly, higher risky behaviors (such as commercial sex activities). Surveys generally find that the proportions of people reporting non-regular sexual relationships are higher in urban areas, although, unsurprisingly, urban behavior also reflects behavior in associated rural areas (Dyson (2003)). Results for the whole sample, sample of women and a sample of men confirm the positive effect of higher population density on HIV contraction, but the effect is not significant in any of the samples.

Overall, adding controls in addition to country-year fixed effects and grid cell fixed effects does not change the estimated effect of submarine Internet on HIV Infection. This effect is remarkably stable across specifications, also implying that the selection of unobserved factors does not drive the effect. As such, these individual-level control variables are not included in the rest of the analyses.

1.6.3 Heterogeneous effects of submarine Internet on HIV

In the preceding section, we studied the "average" effect of submarine Internet on HIV infection, finding that access to this technology helps to reduce exposure to the AIDS virus. In this section, we would like to know if this beneficial effect of the Internet accrues more to

[16] PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables (called principal components). This transformation is defined in such a way that the first principal component accounts for the most significant portion of the variability in the data.

specific groups. To answer this question, we study the heterogeneous effect of Internet access by age, education level, and urban/rural residence. This analysis is likely to inform some of the mechanisms through which the Internet affects sexual behavior and HIV, given that individuals who are more educated, older, and who reside in urban areas have greater Internet access. This analysis also has the potential to inspire policy interventions, allowing policies to be better targeted toward certain groups. This exercise is conducted for all individuals and separately for men and women.

Education level

In columns 2,3, and 4 of table 1-5, we split the sample into three sub-samples representing individuals with primary education, secondary or higher education, and no education, respectively. We estimate the effect of submarine Internet separately for each sub-sample. Panel A shows that access to the submarine Internet reduced HIV by 2.1 percentage points among individuals with secondary or higher education levels and by 2.6 percentage points among those with primary education, but the effect is not statistically significant for this latter group. The effect of submarine Internet on HIV infection is much lower for individuals who have no education (1.6 percentage points), and it is not statistically significant. Panels B and C show that only women drive these effects. Using a computer or absorbing the information coming from the Internet requires that users be able to read and write. However, in contexts where educated and non-educated individuals live together, spillover effects flowing from the former group to the latter are possible, which might explain why the Internet has some effect on non-educated individuals too.

Place of residence

We estimate the effect of submarine Internet separately for urban and rural areas in columns (5) and (6) of table 1-5. The estimates show that the effect is only present in urban areas and is uniquely driven by urban women. Internet access reduces the likelihood of HIV infection in this latter subgroup by 3.1 percentage points, about ten times greater than its effect among rural women. The effect is also close to zero among men in both urban and rural areas. The urban-rural differential effect of submarine Internet can be explained by the fact that in most developing countries, the quality of telecommunication infrastructures is higher in urban areas due to better maintenance. Moreover, even if the supply and quality of the Internet did not vary across areas, one would still expect the demand for, and use of, the Internet to be greater in urban areas due to higher level of education and because of a greater complementary between this technology and urban productive activities. The analysis suggests that the expansion of the Internet in rural locations should be accompanied by complementary policies that help optimize its use in those areas.

Age

Finally, we investigate the differential effect of submarine Internet on HIV infection by age. We split the sample into two age groups: above 24 years old and below 24 years old. The estimates are presented in the last two columns of table 1-5. We find that the primary beneficiaries of submarine Internet are older individuals, especially women older than 24 years old. Access to this technology reduces the likelihood of HIV infection in this latter group by 3.8 percentage points and by only 0.01 percentage points among women younger than 24 years old. The effect is close to zero among men and varies little by age group.

These findings are consistent with the fact that older individuals are richer and more educated, and therefore have a greater ability to access the Internet and make use of its content. They are also consistent with the idea that a significant number of African mothers use the Internet to answer health questions related to their kids, which also exposes them to more information related to HIV. Moreover, the analysis underscores the need for targeting educational programs about HIV and sexual relations among teenagers (see Duflo et al. (2015b), Duflo et al. (2015a), and Dupas (2011)), as our findings imply that they are not the primary beneficiaries of the positive effect of submarine Internet on HIV reduction.

1.7 Mechanisms

This section analyzes several mechanisms that may explain our main finding that connection to submarine Internet cables decreases the likelihood of HIV infection in locations close to the Internet network. Following our conceptual framework in Figure 1-5, We first investigate if access to the Internet affects the Internet's take-up rate among people then test two main channels through which the Internet affects HIV prevalence. These mechanisms consist of the comprehensive knowledge of HIV transmission modes and prevention methods, and sexual behaviors. To test the latter channel, we consider several variables that are interesting in their own rights. These variables are condom use, age at first sexual intercourse, number of sexual partners, marital status, involvement in a polygamous marriage, and fertility choice. We test these channels for the whole sample and separately for women and men.

1.7.1 Internet take-up rate

Following the arrival of submarine cables in the coastal regions of Africa, we expect a higher Internet take-up rate in these regions. We assess the extent to which the arrival of fast

Internet affects an individual's Internet take-up rate using equation 1.1, where the dependent variable is alternatively a dummy variable indicating daily usage of the Internet or usage of the Internet at any time in a month. Information Internet usage comes from rounds 4 and 5 of the Afrobarometer surveys conducted in 8 African coastal countries.

The results are presented in table 1-6. Panel A shows that access to fast Internet increases Internet usage by 3.5 percentage points for daily users and by 4.5 percentage points for any usage during a month. Four of the 8 countries used for the analysis do not have data on HIV. In Panel B, we restrict the analysis to the four remaining countries for which we have data on HIV. These countries are Ghana, Kenya, Mozambique, and Tanzania. The results confirm that submarine cables raise the Internet take-up rate in this group of countries. We expect this effect to increase information on HIV transmission modes and prevention methods.

1.7.2 Comprehensive knowledge of HIV

Accurate knowledge of HIV transmission modes and prevention methods is key to preventing HIV infection. This knowledge is also an essential factor for the success of global initiatives in encouraging the early diagnosis of HIV. In 2001, the United Nations General Assembly Special Session on HIV/AIDS (UNGASS) included HIV knowledge among individuals aged 15-24 years as one of its core indicators of national HIV programmes' success (Gall et al. (2017)). There are two main categories of HIV knowledge: knowledge of how to prevent the disease and knowledge of how the disease is transmitted. These categories are used due to their perceived association with sexual risk and HIV testing behaviors. In sub-Saharan Africa, one of the fundamental ways of preventing the spread of HIV is by increasing knowledge of its transmission. Understanding HIV transmission modes and prevention methods enables sexually active individuals to identify and avoid engaging in behaviors that increase HIV infection risk.

The results of the estimation of the effect of submarine Internet access on HIV knowledge are reported in column 1 of table 1-7. We find that the HIV knowledge index increases by 0.089 points out of 5 following the arrival of the submarine Internet. This effect is statistically significant at a 5 percent confidence interval. A separate analysis by gender reveals that the effect is bigger in the sample of women. Indeed, the HIV knowledge index for women living in an area close to the Internet network increased by 0.136 points due to access to submarine Internet, while for males, it increased by 0.01 points, which is not statistically significant. These results confirm that, since women are more susceptible to HIV in Africa and may face social and cultural barriers to HIV knowledge, they are the primary beneficiary of the technological improvement brought about by submarine cables. This finding has important policy implications, namely that the Internet should be promoted as a source of HIV/AIDS information to help women protect themselves against this disease.

1.7.3 Condom use

The use of condoms is a measure that effectively protects against the transmission of sexually transmitted infections such as HIV. Access to relatively high-speed Internet has a theoretically ambiguous effect on condom use. On the one hand, it is likely to increase exposure to erotic and sexually explicit content, and it is found that frequent consumption of pornography is associated with less consistent use of condoms (Wright et al. (2019)). At the same time, Internet access may increase exposure to health-related information, including information about the effective use of a condom. We report the estimation results of the effect of submarine Internet on condom use in column 2 of table 1-7. We find that condom use during the last sexual intercourse increases by 2.1 percentage points following the arrival of submarine Internet, and women drive this effect. For women in connected locations, the introduction of submarine Internet cables is associated with a 2.3 percentage points increase in the probability of using a condom in their intercourse with their partners. For men, the arrival of submarine Internet increases condom use by 1.3 percentage points, but this effect is not statistically significant. These findings are consistent with the effect of the submarine Internet on modern contraceptive use in general. Column 3 of table 1-7 presents that modern contraception (including contraception that only prevents pregnancy) increases by roughly 2 percentage points following the arrival of submarine Internet. This increase is only observed among women (3 percentage points), as the effect for men is negative, although it is very small and is not statistically significant. The results in column 4 of the table do not show any significant change in traditional contraceptive use after the arrival of submarine cables in coastal regions.

1.7.4 Extramarital relationship

The Internet has the potential to reduce the search cost of sexual partners. Additionally, individuals who find partners via the Internet may be more likely to have more than one sexual partner simultaneously, increasing their likelihood of HIV infection. At the same time, people who use the Internet to search for sexual partners may have a greater probability of finding a suitable match, which can increase relationship stability resulting in less infidelity.

We report the results of the estimation of the effect of access to submarine Internet on the likelihood of an extramarital sexual relationship in column 5 of table 1-7. Due to access to this technology, we find that the probability of having a partner other than the spouse increases by nearly 2 percentage points. The effect is larger for women (+2.8 percentage points) than for men (+1.4 percentage points) and is only statistically significant for the former. The differential effect of submarine Internet for men and women reflects the reality of male dominance in most African societies. Male infidelity is more tolerated than female infidelity (Pongou and Serrano (2013)), so the marginal effect of the Internet on men's sexual behavior should be more negligible. The Internet makes it possible for women to find partners in communities other than their own, allowing them to maintain multiple sexual relationships anonymously.

1.7.5 Polygamy

Male polygamy–the dominant form of polygamy in sub-Saharan African societies–has been shown to lead to sexual networks that bias HIV infection against women (Pongou and Serrano (2013)). Although polygamy has been declining (Fenske (2015); Uzuke and Oyeka (2017)), it remains prevalent in many sub-Saharan African countries, particularly in western Africa. There are many socio-cultural, institutional, economic, and historical factors behind the practice of polygamy in Africa (Anderson (2018); Bertocchi and Dimico (2019); Pongou and Serrano (2013)).

Column 6 of table 1-7 shows that polygamy decreases by 2.7 percentage points following the arrival of submarine Internet. This effect is larger among women (-3.19 percentage points) than men (-1.81 percentage points) and is only statistically significant for the former. The results imply that the introduction of high-speed Internet has changed the attitude towards polygamy in coastal African countries and is an important contributing factor to its positive effect on HIV reduction among women.

1.7.6 Age at first sex

Engaging in sexual interactions at very young ages certainly contributes to the spread of sexually transmitted infections such as HIV. A growing body of literature has found an association between the age at first sex and HIV (Magnusson et al. (2015)). As noted earlier, Internet access increases exposure to sexually explicit movies and websites, raising the demand for sex even among adolescents and resulting in early sexual initiation. We investigate this hypothesis and present the results in column 7 of table 1-7. We find that age at first sex is reduced by less than half a year for the whole sample (0.302 years) following the arrival of submarine Internet. This effect is larger among women. On average, women in the treated locations involve in sexual relations around half a year earlier than those living in the unconnected regions. The findings suggest that online anonymity gives women more freedom to make sexual choices without social backlash. Given that man already have a greater level of sexual freedom even in the absence of the Internet, access to faster Internet with higher capacity does not significantly affect them.

1.7.7 Marriage

Marriage is generally perceived as a protective institution against HIV. It is shown that marriage reduces the number of partners a person might have, with a mitigating effect on HIV (Greenwood et al. (2017)). The Internet has a theoretically ambiguous effect on marriage. On the one hand, it is argued that the Internet suddenly opens a broader choice set of potential partners available online and undermines marriage and other primary relationships (Turkle (2016)). On the other hand, the Internet has the potential to reduce search frictions by allowing individuals to identify a faster and more extensive set of available options that conform to their preferences. Hence it can affect matching and marriage in general.

The estimation of the effect of submarine Internet on marital status is reported in column 8 of table 1-7. We find that access to submarine Internet reduces the likelihood of being married by 1 percentage point, but this effect is not statistically significant. The effect does not differ much between men (-1.21 percentage points) and women (-1.33 percentage points). Even if the effect is not significant, the negative sign seems to imply that the Internet undermines the marriage institution.

1.7.8 Fertility choice

The decision to have a child or to become pregnant can contribute to being infected with HIV. Women who adopt abstinence or condom use in response to HIV prevention programs are less likely to be pregnant and equally less likely to contract HIV infection. Access to the Internet has an ambiguous effect on women's decision to have a child. The Internet provides access to information about contraceptive use and the possible life course consequences of the choice to become a parent (Billari et al. (2019)). Also, it has a fertility-decreasing impact at younger ages through the information mechanism (Guldi and Herbst (2017)). At the same time, Internet may affect the likelihood of finding a partner to become the parent of a joint child (Rosenfeld and Thomas (2012)). This might lead to increased fertility.

We define fertility as the probability of progressing to a second or higher-order child for married women during the last 12 months before the survey. We estimate the effect of submarine Internet on fertility for three age categories. The results are shown in table 1-8. We find that submarine Internet access decreases the fertility of women aged 15–45 years old. The effect is significant for the age group of 25-35. The results align with our findings that access to submarine Internet increases the use of modern contraception and decreases the likelihood of getting married among women.

1.7.9 Migration

We check whether migration is driving the effect of submarine Internet on HIV. Migration could also be a source of bias due to the possible mobility of HIV-infected individuals towards locations close to the Internet network after the arrival of submarine cables. To investigate this issue, we estimate the causal effect of fast Internet among survey respondents who did not change their place of residence during the last ten years before the survey.[17] Results are presented in table 1-9. We find that Internet access significantly decreases HIV prevalence in this restricted sample, which shows that migration is not an issue for our analysis.

1.7.10 Employment status and occupational choice

Recent studies show that access to the fast Internet increases the probability of finding a job in general, and more specifically, a high-skilled job (see, e.g., Hjort and Poulsen (2019), Chiplunkar and Goldberg (2022), Lebrand et al. (2023)). We check if the effect of the submarine on HIV spread in our setting is driven by employment status or type of occupation. It is possible that individuals with a (high-skilled) job are more or less involved in high-risk sexual activities, increasing or decreasing their risk of HIV infection. We show that this is not the case. First, we estimate the effect of submarine Internet on HIV both among individuals who had a job in the last 12 months preceding the survey and those who did not. Second, among individuals who had a job, we estimate the effect of submarine Internet both among those who have high-skilled jobs and those who have low-skilled jobs.[18] The results are presented in table 1-10. We find that following the arrival of submarine Internet, HIV decreases by about 2.9 percentage points among individuals who have a job and by about 2.7 percentage points among individuals who do not have a job. It follows that the magnitude of the effect of fast Internet does not vary much by employment status. Moreover, among employed individuals, access to submarine Internet decreases the risk of HIV infection by 2.3 percentage points among those with high-skilled jobs and by 6.8 percentage points among those with low-skilled jobs. This latter finding demonstrates that occupation type does not

[17] This information is unavailable for all countries in our data set. It is available for the Democratic Republic of the Congo, Kenya, Ghana and Tanzania.

[18] A job is categorized as skilled if it is one of the following categories: professional, clerical, manual-skilled, sales, services, and agricultural employed; a job is categorized as low-skilled if it is one of the following categories: manual-unskilled, domestic, and agricultural self-employed.

explain the effect of the fast Internet on HIV. If anything, the effect is larger for low-skilled workers, which form the employment category that did not benefit from the expansion of fast Internet in Africa (Hjort and Poulsen (2019)). These findings corroborate the fact that our results are robust to controlling for employment status in the regression estimated over the whole sample (see table 1-2).

In summary, it follows from the analysis of operating channels that while access to high-speed Internet increases the likelihood of early sexual initiation and extramarital relationships, it also increases knowledge of HIV prevention and transmission, promotes the use of modern methods of contraception, and reduces fertility and polygamy. Our analysis of Google Trends data corroborates these findings and demonstrates that improved Internet access leads to an increase in the time spent online researching sex-related topics and information on STIs. Importantly, we rule out employment status, occupational choice, and migration as possible channels of impact.

1.7.11 Internet expansion and awareness of HIV and sexual behavior related topics: Evidence from Google trends data

In order to show how the level of people's awareness is changing with Internet penetration, we use Google Trends data for the popularity of a particular search term relative to the maximum volume of search activities as an indicator of searching interest for a specific topic over the years of 2009 to 2019. We then find the correlation between the searching interest and the number of active Internet nodes in Africa during the same period for eleven African coastal countries. We selected 12 relevant words related to our suggestive conceptual framework. Figure 1-8 represents how the search interest of people on HIV and different sexual behavior topics in Google (the red line) changes with the number of Internet nodes (the blue line) between 2009 and 2019. Each fragment represents that as the Internet nodes increase over time, the searching interest of people on HIV/AIDS and other related topics increases as well.

Table 1-A.4 outlines the correlation between the number of Internet nodes and the popularity of search for a specific topic. Column 1 presents the results for the simple panel pool estimates. Adding one node in the region increases the search term for HIV/AIDS by 2 percentage points relative to the maximum volume of search activities. Also, the inquiry about using condoms, marriage, polygamy, and sexually transmitted infections is increased by increasing Internet nodes. At the same, people are looking more for a sexual partner and porn topics as a response to increased Internet nodes. Column 2 controls for the country and year fixed effects. The effects of Internet access are qualitatively similar to those obtained

in Column 1, although the magnitude is smaller. Interestingly, although the search interest for sexual partners increases with the level of Internet penetration, the search interest for marriage decreases. This result is consistent with the idea that the Internet lowers the search cost of sexual partners, and it is also aligned with the negative effect of the submarine Internet on marriage found in our analysis.

1.8 Robustness checks and sensitivity analyses

In this section, we conduct several robustness checks to confirm our main finding that Internet access has causally contributed to reducing HIV incidence in Africa. These analyses investigate possible violations of the identifying assumption of parallel trends and show that this paper's findings are robust. Similar to Hjort and Poulsen (2019), we prove that changing grid cell size and the connectivity radius does not affect the results. Also, we account for the effects of other infrastructures, such as roads and electricity in estimating the effect of connection to submarine Internet on HIV. Further, we use alternative data sets with varying network sizes during 2009-2019 and proximity to the internet network (continuous variable) to confirm our main finding. Finally, we investigate if the Spatial correlation of error terms for limited variation is an issue for the causal relationship.

1.8.1 Change in the connectivity radius

One possible issue with this paper's identification strategy is the radius used to define a location's connectivity status. We show that the estimated effect of submarine Internet on HIV infection is not too sensitive to varying the radius around the backbone network used to define the treatment group. Figure 1-9 presents point estimates with corresponding confidence intervals for a range of radii from the backbone network in separate regressions. For the main outcome variables, results are displayed for different radii, showing the radius beyond which the point estimate becomes statistically insignificant.

We find that slightly changing the connectivity radius does not change the effect of submarine Internet on HIV. As the radius increases further above 2 kilometer from the internet network, this effect tends to decrease and become statistically insignificant. The fact that the effect becomes insignificant beyond a certain distance confirms that the effect of access to the Internet on the outcome is real. This result also provides a placebo test showing that beyond a certain distance to the Internet backbone network, living just a bit closer to, or a bit farther from, this network does not matter.[19] These observations are qualitatively similar when we consider other outcomes, including the HIV knowledge index, polygamy, and the use of modern methods of contraception. As the connectivity radius increases beyond a certain distance, the effect of submarine Internet decreases and becomes statistically insignificant, which again means that having access to high-speed Internet matters.

1.8.2 Changing grid cell size

We also conduct a robustness check in which we change the grid cells' size used to define the location panel. Table 1-11 presents the effect of submarine Internet for different sizes of the grid cells, and our estimates are remarkably stable. This confirms that this paper's identification strategy is not sensitive to the grid cell size. However, we believe that using a relatively small size (10km-by-10km) as we do in our analyses accounts for a wide range of factors that vary across space.

1.8.3 Controlling for electricity and road networks

One may argue that infrastructures are usually provided as a bundle. So our estimated effect of the submarine Internet is possibly capturing the effects of other infrastructures such as electricity and roads or the interactions between submarine Internet of these infrastructures. Using data on roads and electricity provided to us by the World Bank, we construct two variables measuring connectivity to these two infrastructures.[20] Following a similar approach as in Hjort and Poulsen (2019), we regress HIV status on our treatment variable, controlling for electricity and roads as well as two interaction terms between our treatment and these two variables. The results in panel A of table 1-12 represent the coefficient on our treatment variable (-2.41 percentage points) close to the coefficient obtained in table 1-2, where we do not control for the interaction terms included in this specification. Again, this effect is driven by women (-2.77 percentage points). Notably, the coefficients on the interaction terms are very small and statistically insignificant. These add on HIV infection,strengthening the causal nature of this effect.

[19] Indeed, if we define connectivity based on a fake network that is far enough from the real network, submarine Internet will not matter to our outcomes.

[20] An individual is considered connected to electricity (resp. the road network) if the centroid of the DHS cluster in which this individual resides is within 1.5 km of an electricity grid (resp. 1.5 km of the road network.

1.8.4 Using alternative treatment variables

In table 1-13, we define alternative treatment variables to provide further confirmation of its main finding regarding the effect of submarine Internet on HIV infection. We use data on the expansion of Internet nodes from 2009 until 2019 to define a location's connectivity to the Internet network. Our identification strategy exploits within-grid-cell variation in connectivity to the Internet nodes, which is changing over time. Panel A of table 1-13 summarizes the estimated effects of connection to Internet nodes on HIV. We find that the probability of contracting HIV for an individual connected to the Internet nodes decreases by 1.2 percentage points after the arrival of the submarine Internet. We estimate this effect by gender and find that it is entirely driven by women, who experience a 1.8 percentage point decline in HIV infection following the introduction of submarine cables.

In panel B of table 1-13, we estimate how the effect of access to the submarine Internet on HIV changes with distance to the Internet network. In table 1-13 we provide the treatment effect of arrival submarine Internet, which is the multiplication of negative distance to the road based on the kilometre and dummy variable of whether the individual's location is connected to the submarine Internet or not. Being one kilometre closer to the Internet network decreases HIV prevalence by 2.3 percentage points. This effect is larger in the sample of women. Women who are one kilometre closer to the Internet network, are 2.7 percentage points less possible to become infected with HIV.

1.8.5 Spatial correlation of error term when variation is limited

Conley and Taber (2011) point out that results of difference-in-differences estimation with few treated clusters might be biased. They propose wild cluster bootstrap as a particular application of randomization inference. MacKinnon and Webb (2017) show that the wild cluster bootstrap fails in this setting; unrestricted wild cluster over-rejects and restricted wild cluster under-rejects. They propose the subcluster bootstrap to reduce the extent of the problem. Given that in this study, we are using a limited variation in the timing of the arrival of submarine cables in Africa, the cluster-robust results can be biased. We perform variants of the wild cluster bootstrap to test the hypothesis that the arrival of submarine cables was not associated with the HIV outbreak. Table 1-A.5 presents parameter estimates, cluster-robust standard errors at the grid 10 km $\times$ 10 km cells level, and t-statistics for the effect of submarine Internet on HIV prevalence, along with eight bootstrap P-values and the associated 95% confidence intervals. Four tests impose the null hypothesis, and four do not. Within each triplet, the first variant clusters the bootstrap errors by grid 10 km $\times$ 10 km cells, the level at which they are clustered in our specifications. The second and third variants cluster the bootstrap errors by grid 10 km $\times$ 10 km cells, country and year combination, allowing for arbitrarily correlated error terms within the same country or year. The fourth variant clusters the bootstrap at the individual level, which is the whole sample, and this variant is the ordinary wild bootstrap. Our results are robust across all variants, implying that the limited number of treated grid cells is not an issue for our analysis.

1.9 Conclusion

The HIV/AIDS epidemic remains one of the most severe public health crises of our time. In this paper, we have argued that high-speed Internet, as an information and communication technology, is likely to affect the distal and proximate causes of HIV spread in non-trivial ways. In so doing, we have provided the first evidence of the causal impact of access to submarine Internet on HIV spread in Africa. Our research design utilizes the gradual arrival of submarine cables from Europe in Africa's coastal countries between 2009-2012 and follows a difference-in-differences strategy for identification. Our main finding is that HIV incidence decreases by over 2 percentage points among individuals connected to the Internet backbones following the arrival of submarine cables. This beneficial effect of submarine Internet only accrues to women. Among women, we also uncover a great deal of heterogeneity in the impact of this technological improvement. Women living in urban areas, educated women, and women over 24 years old are the primary beneficiaries. These findings are consistent with the idea that a minimal level of education and financial means is needed to reap the full benefits of the Internet. They underscore the need for targeted HIV prevention programs for younger and disadvantaged women for maximal efficiency.

Analyzing the underlying mechanisms of the effect of submarine Internet on HIV spread, we uncovered two opposing effects of this technology using both individual-level data and Google Trends data. On the one hand, we found that access to the submarine Internet increases knowledge of HIV transmission modes and prevention methods, increases the use of condoms and other modern methods of contraception, and decreases polygamy and fertility. On the other hand, we found that the introduction of submarine Internet lowers the age of the first sex, decreases the likelihood of getting married, and encourages extramarital relationships. These latter findings are consistent with the view that the Internet lowers the cost of searching for sexual partners, and our analysis of the Google trends data also corroborates them. Indeed, combining data from Google Trends with data on the growth of Internet nodes over the last decade, we find that increased Internet access boots the online search popularity of sex-related topics (having sex, sexual words, pornography, sexual abuse), sexual partners, contraception, HIV/AIDS, and other sexually transmitted diseases.

Importantly, the fact that submarine Internet significantly decreases HIV incidence implies that its positive effects on some of the proximate causes of this virus outweigh its negative effects. Also, the fact that the effect of this technology is only present to women implies that the expansion of the fast Internet is reducing the gender gap in HIV prevalence in Africa, which is the only continent where HIV rates are much higher in women than men.

The paper's findings suggest that expanding high-speed Internet infrastructure may be among the most feasible policy options to control HIV in Africa and may also improve the efficacy of community-based programs. However, the negative impact of this technology on certain family outcomes means that its expansion should be accompanied by other educational policies that optimize its utilisation.

Figure 1-1: Internet nodes in Africa in 2009 and 2019

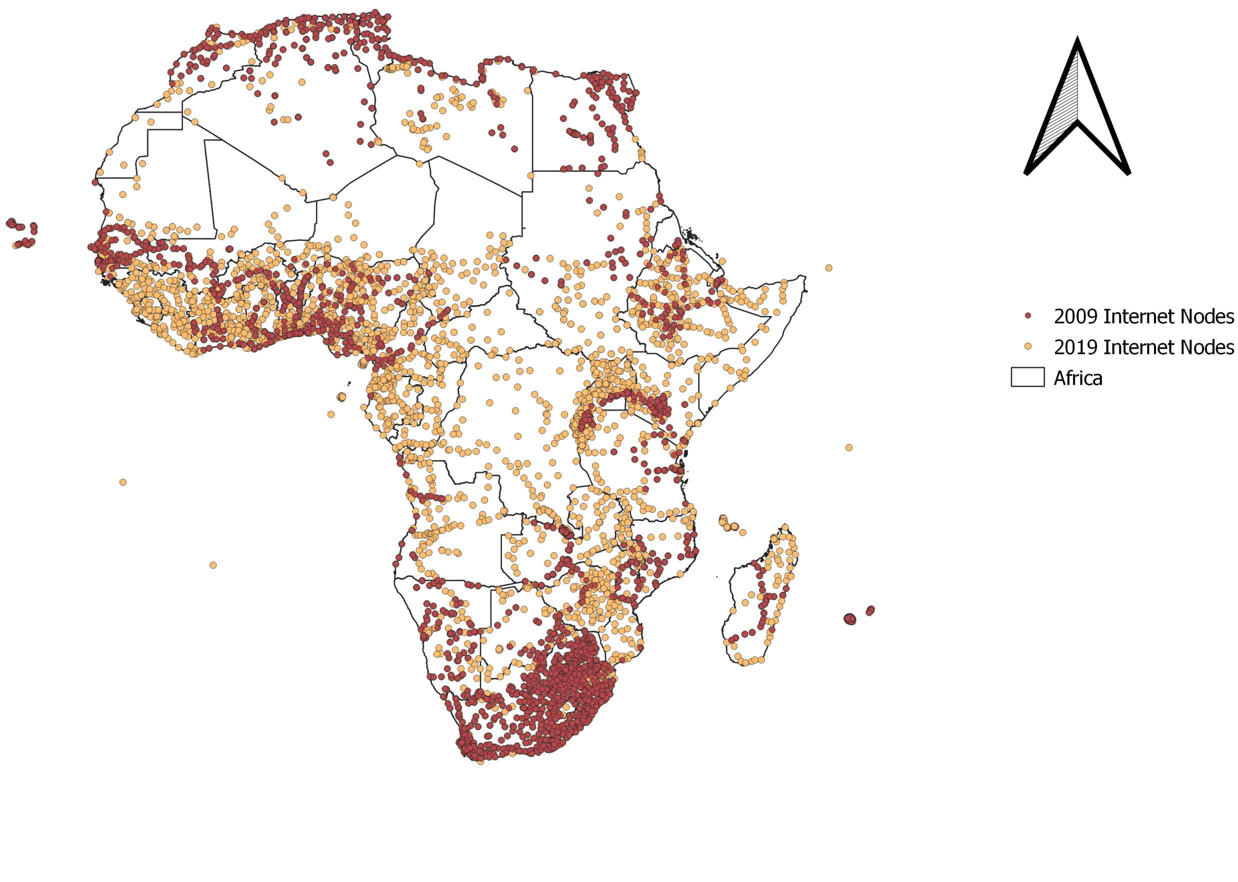


Figure 1-2: Evolution of the number of Internet nodes between 2009-2019

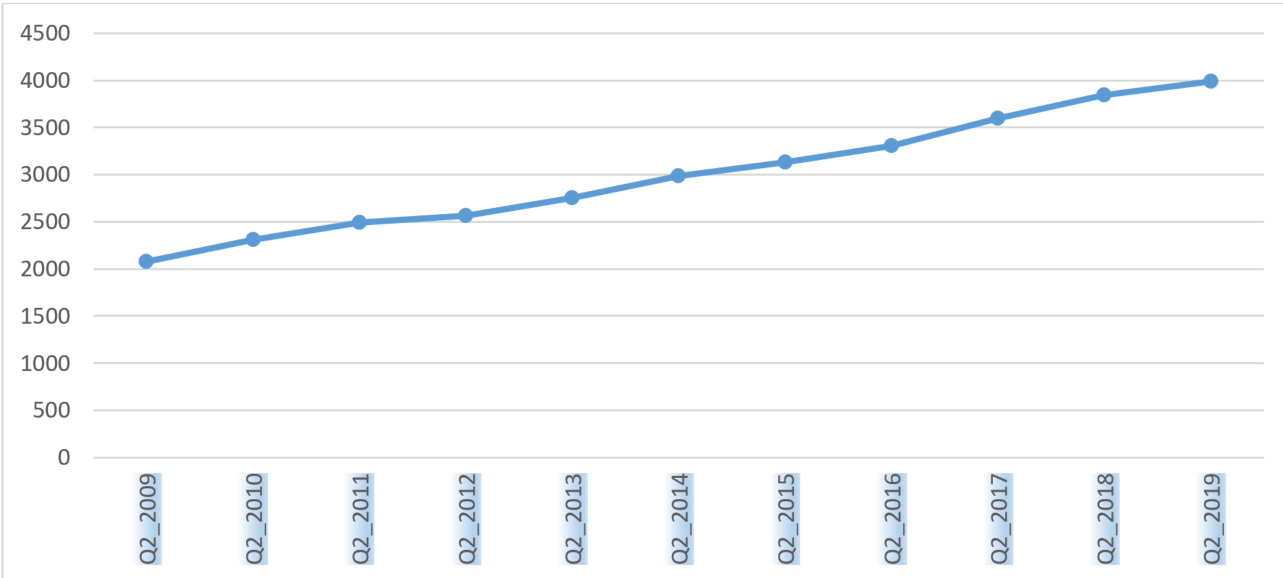


Figure 1-3: Internet backbone network in 2009

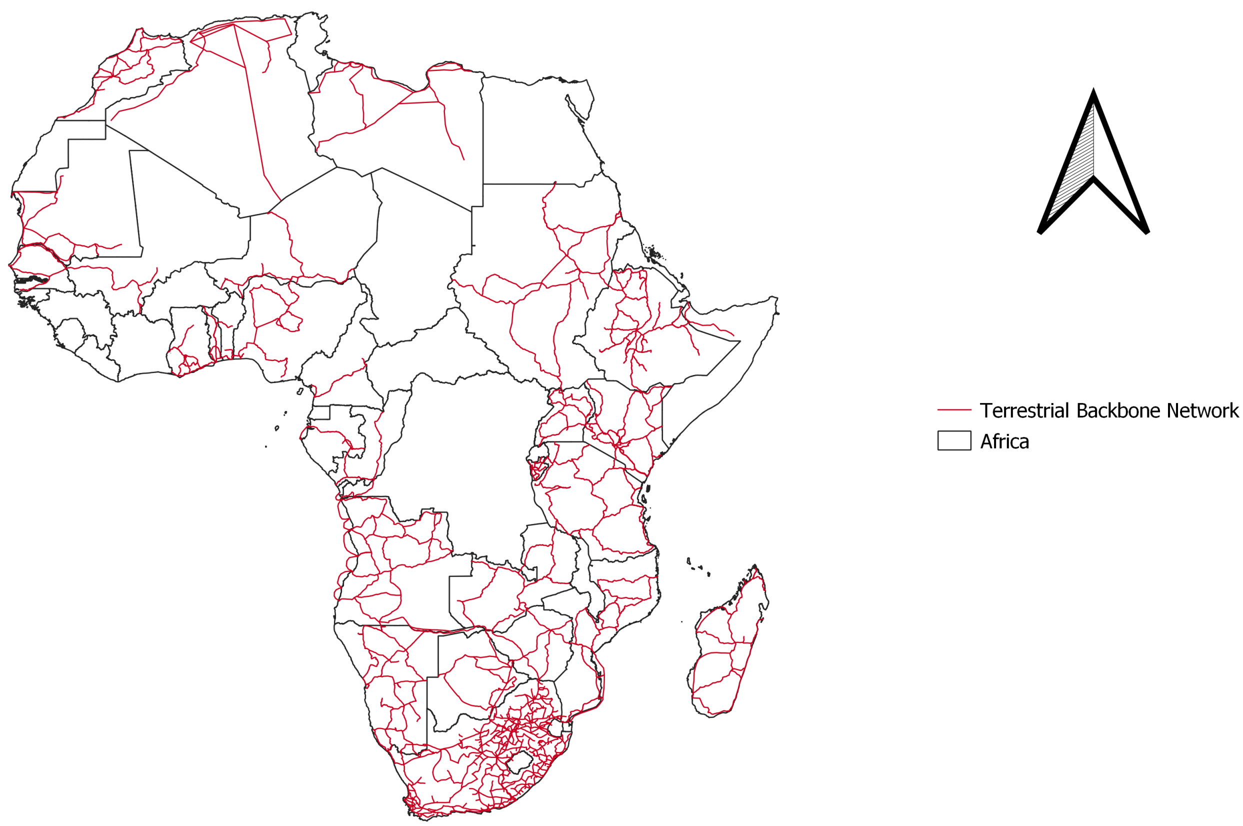


Figure 1-4: Arrival of submarine cables between 2009 and 2012

(a) 2009

(b) 2010

(c) 2011

(d) 2012

Source: Submarine Cable Map (https://www.submarinecablemap.com).

Figure 1-5: Conceptual framework for the effect of fast Internet on HIV

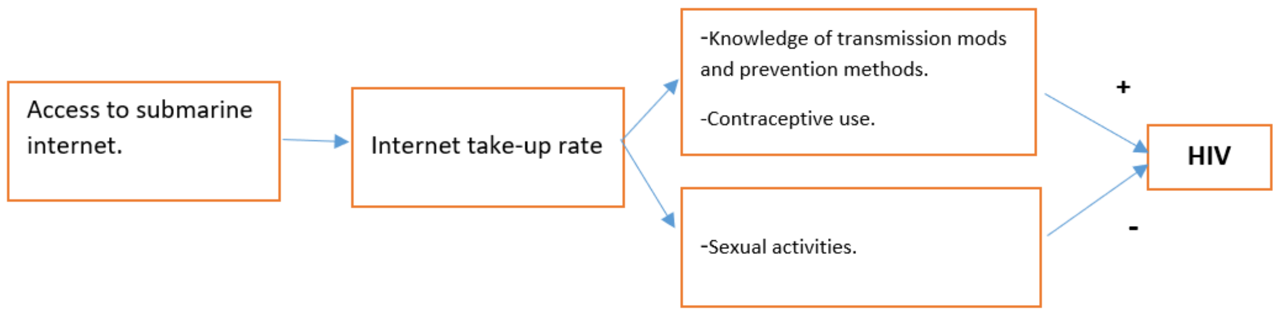
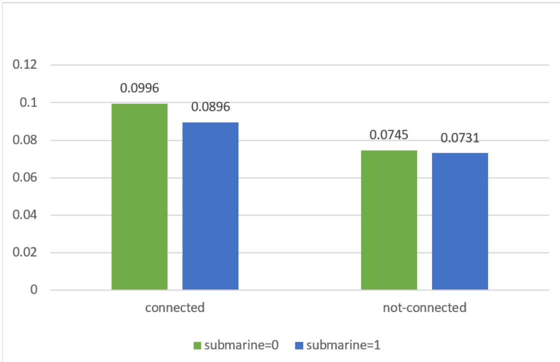
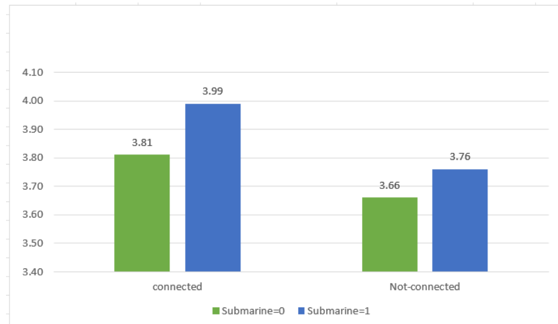


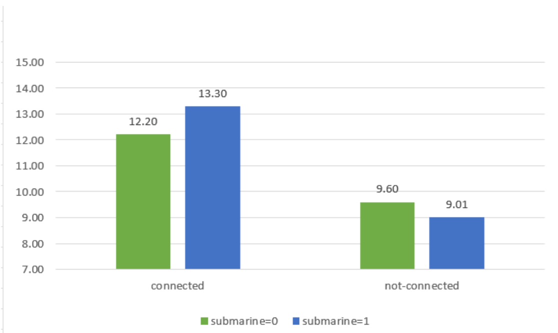
Figure 1-6: Change in outcome variables before and after arrival of fast Internet



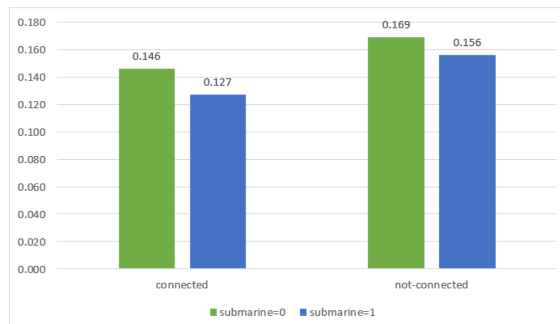
(a) HIV-infected.



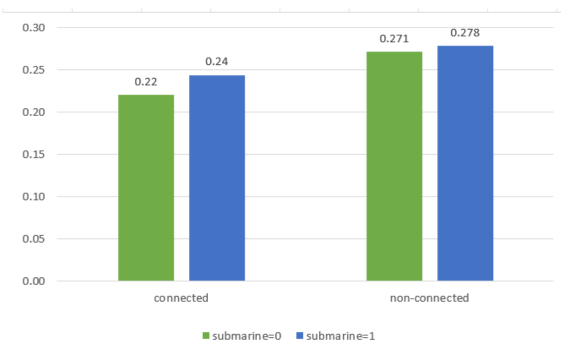
(b) HIV knowledge index.



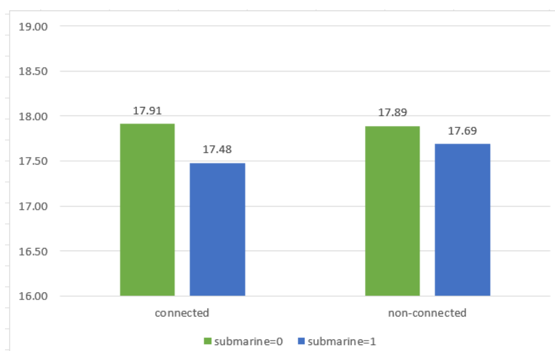
(c) Condom use.



(d) Involved in a polygamous marriage.



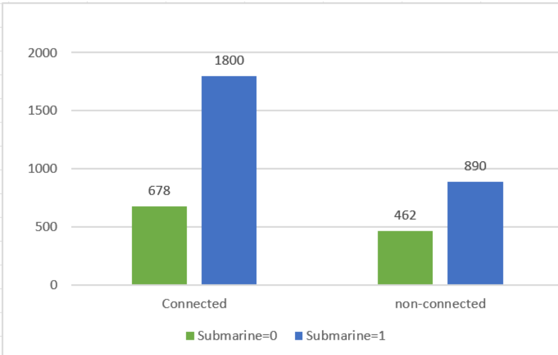
(e) Have partners other than the spouse.



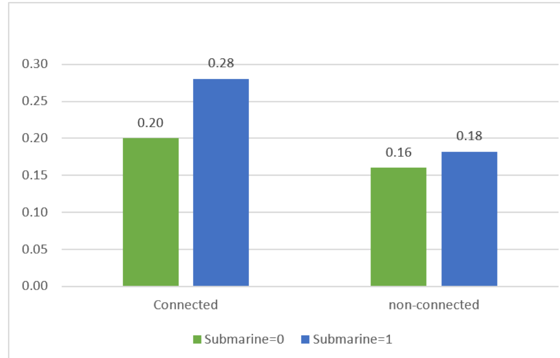
(f) Age at first intercourse.

Notes: Data for Internet Akamai content delivery come from Hjort and Poulsen (2019). Hjort and Poulsen (2019) use quarterly data on average connection speeds for 900 African locations during the 2007–2013 period. Hjort and Poulsen (2019) restrict to location-quarters for which the speed measure is based on more than ten unique IP addresses. Information on Internet take-up rates comes from the Afrobarometer surveys. Countries and years are listed in table 1-A.1.

Figure 1-7: Change in Internet speed and take-up rate before and after arrival of fast Internet

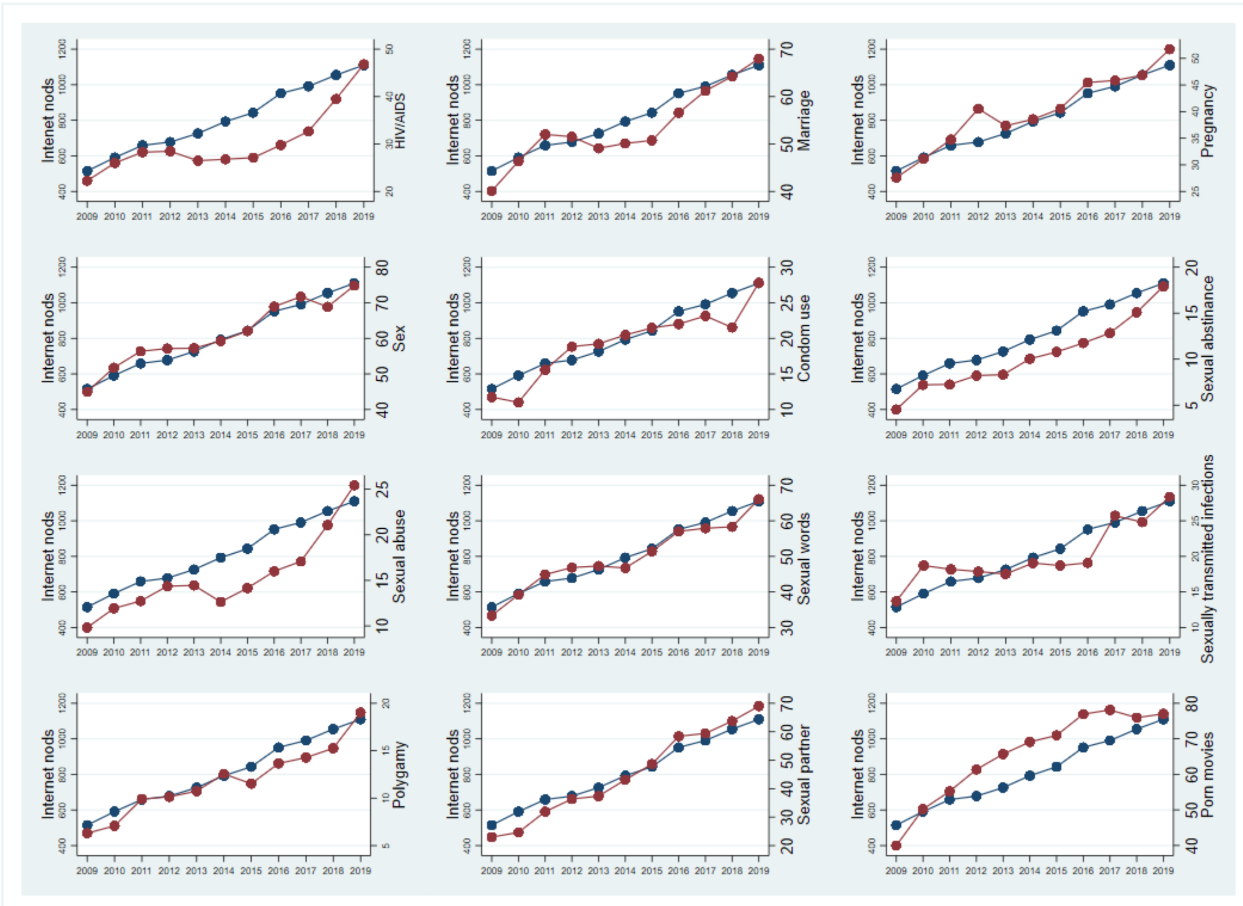


(a) Internet speed.



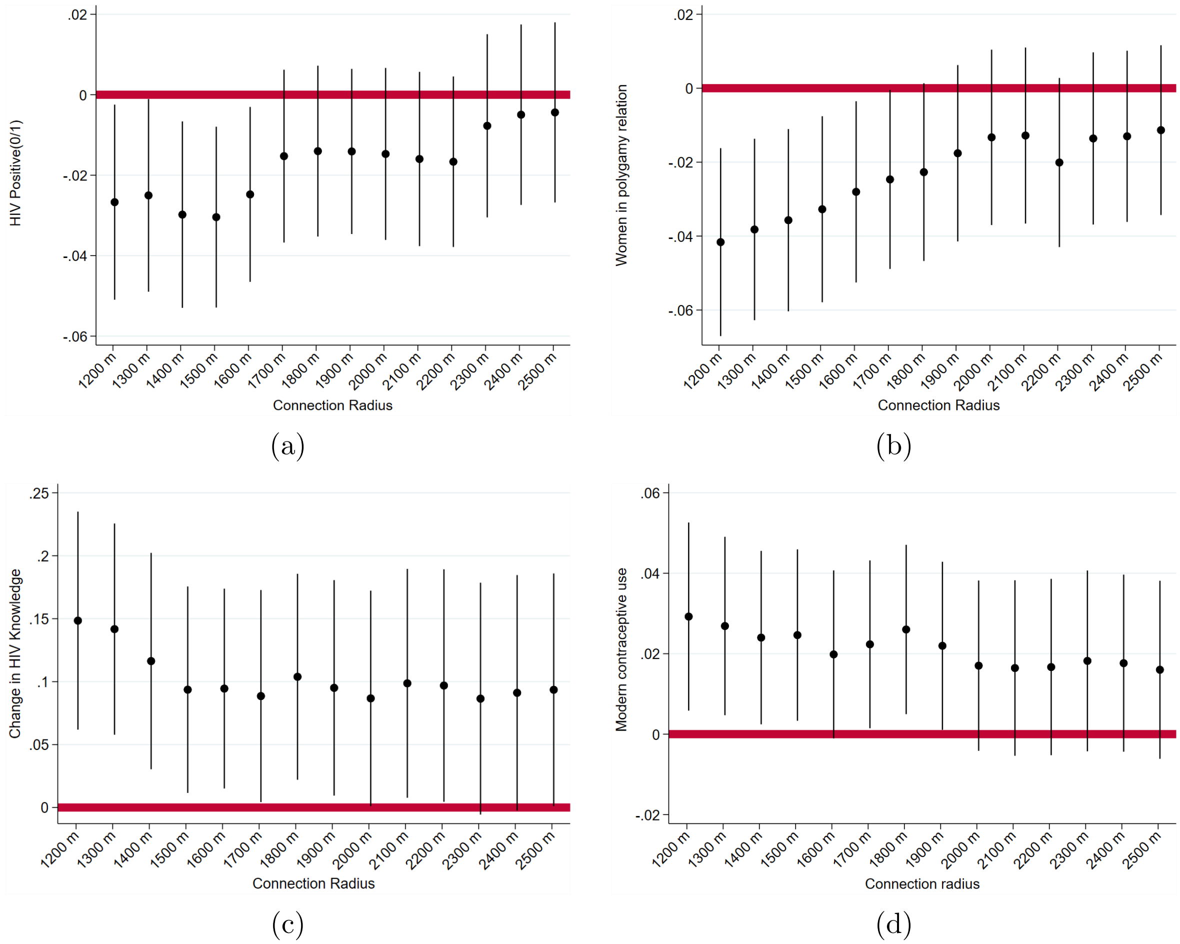
(b) Internet take-up rate.

Figure 1-8: Trends in the number of Internet nodes and search for specific topics between 2009 and 2019



Notes: The blue graphs represent trends in the number of Internet nodes between 2009 and 2019. Each red graph represents trends in the online search score of a corresponding topic between 2009 and 2019; data are from Google Trends.

Figure 1-9: Sensitivity of the point estimate of the effect of submarine Internet on some outcomes to the Internet connection radius



(a) (b) (c) (d)

Table 1-1: Socio-economic characteristics and sexual behavior before the arrival of submarine Internet

Variables	Connected	Not-Connected	Difference
Panel (A): Selected Control Variables			
Age	29.17	29.59	0.42
	(10.78)	(10.92)	(1.32)
Education	0.89	0.86	-0.03
	(0.30)	(0.33)	(-2.13)
Have a job	0.73	0 .76	0.02
	(0.44)	(0.42)	(1.22)
Muslem	0.28	0.30	0.01
	(0.450)	(0.458)	(1.87)
Christian	0.29	0.25	-0.03
	(0.455)	(0.43)	(-2.98)
Exposure to the media	0.95	0.92	-0.03
	(0.20)	(0..26)	(-6.43)
Panel (B): Selected Dependent Variables			
Internet take-up rate	0.23	0.21	-0.02
	(0.42)	(0.41)	(-2.07)
HIV positive	0.10	0.07	-0.03
	(0.30)	(0.26)	(-3.03)
HIV information	3.8	3.6	-0.12
	(1.695)	(1.595)	(-1.86)
Modern contraceptive use	0.28	0.26	-0.02
	(0.45)	(0.43)	(-3.56)
Traditional contraceptive use	0.059	0.054	-0.005
	(0.55)	(0.54)	(-1.89)
Using condom	0.12	0.12	-0.018
	(0.304)	(0.275)	(-4.56)
Polygamy	0.169	0.147	0.0192
	(0.347)	(0.328)	(2.68)
Non-spousal sexual partners	0.22	0.27	-0.05
	(0.374)	(0.408)	(-1.61)
Age at first sex	17.91	17.89	-0.02
	(16.53)	(16.62)	(-1.14)
Married	0.53	0.49	0.038
	(0.496)	(0.491)	(3.40)
Fertility choice rate	0.108	0.123	0.01
	(0.311)	(0.329)	(1.95)

Notes: All statistics presented are from the period before the arrival of submarine cables. Information on Internet speed is in Akamai Quarterly data and comes from Hjort and Poulsen (2019). As in Hjort and Poulsen (2019), we restrict the sample to location-quarters for which the speed is measured using more than ten unique IP addresses; we exclude the four biggest cities in each country. Individuals in the connected locations live within 1500 meters of the backbone network, and individuals in non-connected locations live further. We drop individuals who live >10 Km away. Statistics in parentheses are standard errors in the first three columns and t-statistics in the last column.

Table 1-2: Effect of submarine Internet on HIV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connected-network*submarine	-0.0227**	-0.0238**	-0.0239**	-0.0228**	-0.0228**	-0.0211*	-0.0267**	-0.0269**
	(0.0111)	(0.0110)	(0.0110)	(0.0108)	(0.0108)	(0.0129)	(0.0140)	(0.0141)
Age		0.00237***	0.00226***	0.00218***	0.00218***	0.00215***	0.00222***	0.00221***
		(0.000203)	(0.000205)	(0.000184)	(0.000184)	(0.000203)	(0.000209)	(0.000209)
Have job			-0.00708*	-0.00593*	-0.00631*	-0.00833*	-0.00789*	-0.00781*
			(0.00302)	(0.00299)	(0.00298)	(0.00330)	(0.00328)	(0.00325)
Primary education				0.0105*	0.0117*	0.0161**	0.0165**	0.0163**
				(0.00493)	(0.00498)	(0.00562)	(0.00564)	(0.00561)
Secondary education				-0.0164**	-0.0147*	-0.0124*	-0.0108	-0.0109
				(0.00587)	(0.00582)	(0.00604)	(0.00584)	(0.00585)
High education				-0.0444***	-0.0426***	-0.0400***	-0.0342***	-0.0341***
				(0.00828)	(0.00820)	(0.00823)	(0.00838)	(0.00836)
Exposure to the media					-0.0106	-0.0123*	-0.00996	-0.0102
					(0.00542)	(0.00582)	(0.00584)	(0.00584)
Muslim						0.0000952	0.000101	0.000157
						(0.00595)	(0.00597)	(0.00591)
Christian						-0.00448	-0.00441	-0.00443
						(0.00585)	(0.00582)	(0.00581)
Wealth							-0.0119***	-0.0118***
							(0.00318)	(0.00311)
Population density								0.00000105
								(0.000000986)
NOB	38,527	38,527	38,526	38,526	38,526	32,516	31,944	31,944
R2	0.087	0.098	0.098	0.101	0.101	0.104	0.105	0.105
Country-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Grid-connect FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cells are 0.1 × 0.1 decimal degrees, which is roughly 10 × 10 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.01, *** p<0.001.

Table 1-3: Effect of submarine Internet on HIV (sample of women)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connected-network*submarine	-0.0275**	-0.0282**	-0.0281**	-0.0274**	-0.0276**	-0.0267*	-0.0336**	-0.0337**
	(0.0122)	(0.0122)	(0.0123)	(0.0120)	(0.0121)	(0.0153)	(0.0168)	(0.0169)
Age		0.00243***	0.00216***	0.00203***	0.00202***	0.00195***	0.00202***	0.00202***
		(0.000212)	(0.000222)	(0.000199)	(0.000200)	(0.000227)	(0.000237)	(0.000237)
Have job			-0.0180***	-0.0165***	-0.0168***	-0.0217***	-0.0215***	-0.0214***
			(0.00381)	(0.00382)	(0.00381)	(0.00437)	(0.00435)	(0.00434)
Primary education				0.00748	0.00880	0.0138*	0.0142*	0.0141*
				(0.00558)	(0.00564)	(0.00658)	(0.00663)	(0.00661)
Secondary education				-0.0176*	-0.0157*	-0.0137	-0.0118	-0.0119
				(0.00685)	(0.00685)	(0.00740)	(0.00712)	(0.00715)
High education				-0.0548***	-0.0528***	-0.0517***	-0.0448***	-0.0447***
				(0.0111)	(0.0111)	(0.0117)	(0.0124)	(0.0124)
Exposure to the media					-0.0128*	-0.0151*	-0.0116	-0.0119
					(0.00647)	(0.00713)	(0.00706)	(0.00705)
Muslim						-0.00266	-0.00291	-0.00286
						(0.00695)	(0.00698)	(0.00692)
Christian						-0.00290	-0.00310	-0.00321
						(0.00774)	(0.00764)	(0.00765)
Wealth							-0.0144***	-0.0142***
							(0.00384)	(0.00376)
Population density								0.00000106
								(0.00000115)
N	28,856	28,856	28,855	28,855	28,855	22,848	22,544	22,544
R2	0.092	0.101	0.102	0.105	0.105	0.106	0.107	0.107
Country-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Grid-connect FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cells are 0.1 × 0.1 decimal degrees, which is roughly 10 × 10 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.01, *** p<0.001.

Table 1-4: Effect of submarine Internet on HIV (sample of men)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connected-network*submarine	-0.000747	-0.00704	-0.00712	-0.00685	-0.00668	-0.00695	-0.00999	-0.0105
	(0.0115)	(0.0112)	(0.0112)	(0.0110)	(0.0109)	(0.0108)	(0.0123)	(0.0125)
Age		0.00228***	0.00225***	0.00235***	0.00233***	0.00234***	0.00241***	0.00240***
		(0.000283)	(0.000296)	(0.000297)	(0.000297)	(0.000296)	(0.000301)	(0.000301)
Have job			0.00201	0.00131	0.000984	0.000787	0.0000534	0.0000934
			(0.00501)	(0.00497)	(0.00496)	(0.00497)	(0.00509)	(0.00509)
Primary education				0.0403***	0.0389***	0.0396***	0.0403***	0.0401***
				(0.00902)	(0.00899)	(0.00903)	(0.00911)	(0.00909)
Secondary education				0.0154	0.0130	0.0141	0.0156	0.0154
				(0.00849)	(0.00833)	(0.00826)	(0.00834)	(0.00836)
High education				0.00206	-0.000321	0.00120	0.00529	0.00522
				(0.00973)	(0.00970)	(0.00975)	(0.00982)	(0.00982)
Exposure to the media					0.0222*	0.0222*	0.0231*	0.0231*
					(0.0108)	(0.0108)	(0.0111)	(0.0111)
Muslim						-0.00416	-0.00377	-0.00372
						(0.00738)	(0.00735)	(0.00732)
Christian						-0.00870	-0.00824	-0.00815
						(0.00596)	(0.00601)	(0.00602)
Wealth							-0.00679*	-0.00674*
							(0.00284)	(0.00279)
Population density								0.000000723
								(0.000000878)
N	9,671	9,671	9,671	9,671	9,671	9,668	9,400	9,400
R2	0.114	0.132	0.132	0.136	0.137	0.137	0.137	0.138
Country-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Grid-connect FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cells are 0.1 × 0.1 decimal degrees, which is roughly 10 × 10 km.An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * $p<0.1$, ** $p<0.01$, *** $p<0.001$.

Table 1-5: Heterogeneous effects of submarine Internet on HIV

	Baseline	Primary edu	Higher edu	Not edu	Rural	Urban	Age> 24	*Age* < 24
				Panel (A): Whole sample				
Connected-network*submarine	-0.0227**	-0.0263	-0.0207*	-0.0164	-0.0237	-0.0269*	-0.0399**	0.00250
	(0.0111)	(0.0195)	(0.0124)	(0.0423)	(0.0332)	(0.0137)	-0.0154	(0.0109)
NOB	38,527	18,390	15,719	4,418	16,820	21,707	21,662	16,865
R2	0.09	0.12	0.09	0.30	0.11	0.08	0.13	0.09
				Panel (B): Sample of women				
Connected-network*submarine	-0.0275**	-0.0224	-0.0378**	-0.0181	-0.0248	-0.0318**	-0.0505**	0.00424
	(0.0122)	(0.0222)	(0.0167)	(0.0468)	(0.0426)	(0.0146)	(0.0170)	(0.0124)
NOB	28,856	15,190	9,901	3,765	12,819	16,037	16,194	12,662
R2	0.09	0.12	0.11	0.32	0.12	0.08	0.14	0.11
				Panel (C): Sample of men				
Connected-network *submarine	-0.000747	0.00813	0.00162	0	-0.00360	-0.00417	-0.0138	0.0130
	(0.0115)	(0.0432)	(0.0102)	(9.75e-18)	(0.0175)	(0.0129)	(0.0177)	(0.00940)
NOB	9,671	3,200	5,818	653	4,001	5,670	5,468	4,203
R2	0.11	0.21	0.14	0.58	0.14	0.11	0.20	0.20
Country-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Grid-connect FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cells are 0.1×0.1 decimal degrees, which is roughly 10×10 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.01, *** p<0.001.

Table 1-6: Effect of submarine Internet on Internet use

	Panel (A)		Panel (B)	
	Daily use	Monthly use	Daily use	Monthly use
Connected-network *submarine	0.0354**	0.0456**	0.0478**	0.0501**
	(0.0131)	(0.0187)	(0.0205)	(0.0246)
NOB	12,555	12,555	5,724	5,724
Country-year FE	✓	✓	✓	✓
Gride-connect FE	✓	✓	✓	✓

Notes: Data for Internet use are from rounds 4 and 5 of Afrobarometer surveys for 8 African countries (see table 1-A.1). Internet use is a dichotomous variable equal to one if the individual uses the Internet daily in the first column and during the year in the second column. An individual's location is considered connected to the Internet network if that location is within 1.5 km of an Internet node.Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.05, *** p<0.001.

Table 1-7: Mechanisms: Effect of submarine Internet on knowledge of HIV and sexual behaviors

	HIV knowledge	Use Condom	Modern Contraceptive	Traditional contraceptive	Non-spousal partner	Polygamy Relations	Age at first intercourse	Marriage
	Whole sample							
Connected*submarine	0.086**	0.021**	0.019*	0.003	0.0215*	-0.0268**	-0.302**	-0.0114
	(0.0429)	(0.00822)	(0.0105)	(0.00502)	(0.0118)	(0.0123)	(0.111)	(0.0111)
NOB	97,293	97,293	96,425	96,425	73,975	46,965	78,324	97,289
	Sample of women							
Connected*submarine	0.135**	0.023**	0.03**	0.004	0.0285**	-0.0319**	-0.458***	-0.0133
	(0.0541)	(0.00969)	(0.012)	(0.00604)	(0.0139)	(0.0147)	(0.126)	(0.0123)
NOB	68,481	68,481	67,989	67,989	50,622	33,377	55,131	68,478
	Sample of men							
Connected*submarine	0.013	0.0139	-0.002	-0.001	0.0145	-0.0181	0.0152	-0.0121
	(0.0555)	(0.0159)	(0.0189)	(0.00990)	(0.0179)	(0.0132)	(0.195)	(0.0158)
NOB	28,812	28,812	28,436	28,436	23,353	13,588	23,193	28,811
Country-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Gride-connect FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cells are 0.1 × 0.1 decimal degrees, which is roughly 10 × 10 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet backbone and a dummy indicator for whether submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.05, *** p<0.001.

Table 1-8: Effect of submarine Internet on fertility

	Whole sample	Age<=25	25<Age<35	35<Age<45
Connected-network*submarine	-0.0171**	-0.0119	-0.0380**	-0.00537
	(0.00837)	(0.0125)	(0.0182)	(0.00914)
NOB	58,054	23,930	18,674	15,450
Country-year FE	✓	✓	✓	✓
Grid-Connect FE	✓	✓	✓	✓

Notes: Fertility is defined as a binary indicator for whether a woman had at least one child or was pregnant during the last 12 months preceding the survey. Grid-cells are 0.1×0.1 decimal degrees, which is roughly 10×10 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 1-9: Effect of the Internet on HIV: controlling for individual's mobility

	Whole sample	Women	Men
Connected-network*submarine	-0.0269**	-0.0336**	0.00585
	(0.0135)	(0.0113)	(0.0182)
NOB	31,662	23,957	7,705
Country-year FE	✓	✓	✓
Grid-line FE	✓	✓	✓

Notes: We control for an individual's mobility in this table. DHS surveys have data on years that individuals resided in the place where they live at the time of the survey for limited countries. In our sample, this information is available only for Congo, Kenya, Ghana and Tanzania. We considered an individual as a mobile if he/she changed the place of residency during the last ten years before the survey. Grid cells are 0.1×0.1 decimal degrees, roughly 10×10 km. Robust standard errors in parentheses are clustered at the grid cell level.* p<0.1, ** p<0.05, *** p<0.001.

Table 1-10: Effect of the Internet on HIV based on employment status and individual's occupation

Panel(A):Effect of Internet on HIV based on employment status			
	whole sample	Have job	Not have job
Connected-network*submarine	-0.0296**	-0.0287**	-0.0267*
	(0.00959)	(0.0102)	(0.0149)
NOB	108,092	82,304	25,788
Panel(B):Effect of the Internet on HIV based on occupation			
		Have skilled job	Have not skilled job
Connected-network*submarine		-0.0225**	-0.0680**
		(0.0111)	(0.0245)
NOB		47,366	33,935
Country-year FE	✓	✓	✓
Grid-line FE	✓	✓	✓

Notes: Grid cells are 0.1×0.1 decimal degrees, which is roughly 10×10 km. An individual's location is considered connected to the Internet network if it is within 1.5 km of the network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. In panel A we split the sample into individuals who had a job during the last 12 months before the survey and those who did not, and in panel B we restrict the analysis to individuals who had a job and split this subsample into individuals who had skilled jobs such as professional, clerical, sales, agricultural self-employed, and skilled manual jobs, and individuals who had low-skilled jobs like manual unskilled, domestic, and agriculture employee. We did not restrict the sample to locations within 10 kilometres of the Internet network for these two panels. Robust standard errors in parentheses are clustered at the grid cell level. * p<0.1, ** p<0.05, *** p<0.001.

Table 1-11: Sensitivity analysis: Effect of submarine Internet on HIV and grid cell size

Grid cell size	10 km	15 km	20 km	25 km	30 km	37 km	40 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Connected-network*submarine	-0.0227**	-0.0200**	-0.0288**	-0.0227**	-0.0263**	-0.0230**	-0.0268**
	(0.0111)	(0.00888)	(0.00968)	(0.0107)	(0.00880)	(0.00856)	(0.00839)
NOB	38,525	38,525	38,525	38,525	38,525	38,525	38,525
Country-year FE	✓	✓	✓	✓	✓	✓	✓
Grid-connect FE	✓	✓	✓	✓	✓	✓	✓

Notes: Grid cell size ranges from 10×10 km to 40×40 km. An individual's location is considered connected if it is within 1.5 km of the backbone network. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the corresponding grid cell level. * p<0.1, ** p<0.05, *** p<0.001.

Table 1-12: Effect of internet on HIV: Controlling for access to the road and electricity

	Whole sample	Women	Men
Connected-network*submarine	-0.0241**	-0.0277**	-0.00373
	(0.0107)	(0.0117)	(0.0113)
Connected-network*submarine* connected-electricity	-0.00404	-0.000506	-0.00651
	(0.00654)	(0.00697)	(0.0106)
Connected-network*submarine* connected-road	0.00733	0.000937	0.0192*
	(0.00735)	(0.00955)	(0.00887)
NOB	38,525	28,854	9,671
Country-year FE	✓	✓	✓
Grid-line FE	✓	✓	✓

Notes: In order to check if the effect of the Internet on HIV is driven by access to other infrastructures, we control for the effect of connectivity to the road and electricity networks. We merged DHS geo-data with the GIS shape files for African electricity grids and roads from the World Bank. An individual's location is considered connected to the backbone network, road network, or electricity if it is within 1.5 km of these networks. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Grid cells are 0.1×0.1 decimal degrees, roughly 10×10 km. Robust standard errors in parentheses are clustered at the grid cell level.* p<0.1, ** p<0.05, *** p<0.001.

Table 1-13: Effect of submarine Internet on HIV (alternative independent variables)

	Whole sample	Women	Men
Panel (A): Connectivity to the internet nodes			
Connected-Internet node*submarine	-0.012*	-0.0180**	-0.00802
	(0.0101)	(0.0081)	(0.0115)
NOB	107,117	66,644	30,503
Panel(B): Proximity to the internet network			
Proximity to Internet node*submarine	-0.0233**	-0.0399*	-0.00720
	(0.0135)	(0.0110)	(0.0147)
NOB	85,635	64,175	21,460
Country-year FE	✓	✓	✓
Grid-line FE	✓	✓	✓

Notes: Grid cells are 0.1×0.1 decimal degrees, which is roughly 10×10 km. The treatment variable is an interaction term between the negative distance (of an individual's location) to the 2009 backbone network and a dummy indicator that submarine cables are available in the survey year. In panel B, an individual's location is considered connected to the Internet if it is within 1.5 km of an Internet node in the survey year. The treatment variable is an interaction term between connectivity to the Internet and a dummy indicator that submarine cables are available in the survey year. Robust standard errors in parentheses are clustered at the grid cell level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

1-A Appendix

Table 1-A.1: Surveys used and date of arrival of submarine cables by country

Country	Submarine cables arrival date	DHS		Afrobarometer
		AIDS Indicator Surveys (AIS)	Standard and geo-coded data	Internet take-up rate
Benin	2010-Q3		(2001, 2012)	(2008, 2012)
Congo	2012-Q2	(2007, 2013)	(2007, 2013)	
Ghana	2010-Q3	(2003, 2014)	(2008, 2014)	(2008, 2012)
Kenya	2009-Q3	(2003, 2010)	(2003, 2014)	(2008, 2012)
Madagascar	2010-Q3			(2008, 2012)
Mozambique	2009-Q3	(2008, 2015)		(2008, 2012)
Namibia	2012-Q2		(2006, 2012)	
Nigeria	2010-Q3		(2008, 2013)	(2008, 2012)
Senegal	2010-Q3			
South Africa	2009-Q3			(2008, 2011)
Tanzania	2009-Q3	(2003, 2011)	(1999, 2010)	(2008, 2012)
Togo	2010-Q3		(1998, 2013)	

Table 1-A.2: Descriptive statistics

Socio-economic and HIV status : AIDS Indicator Surveys(AIS)				
	NOB	Whole	Women	Men
Connected-network * submarine	38,527	0.188	0.182	0.208
Connected-road * submarine	38,527	0.361	0.331	0.451
Connected-Internet node * submarine	38,527	0.057	0.056	0.060
Connected-road * submarine	38,527	0.361	0.331	0.451
Connected- electricity* submarine	38,527	0.199	40.218	0.241
Sex	38,527		0.749	0.251
Age	38,527	29.500	29.281	30.154
High level of education	38,527	0.054	0.035	0.112
low level of education	38,527	0.832	0.835	0.821
No education	38,527	0.114	0.130	0.068
Living in rural	38,527	0.437	0.444	0.414
Living in urban	38,527	0.563	0.556	0.586
Exposure to the media	38,527	0.908	0.901	0.928
Have job	38,527	0.730	0.712	0.785
Muslim	38,527	0.256	0.263	0.242
Christian	38,527	0.239	0.252	0.210
population density(per squre KM)	38,527	3400.227	3233.314	3898.255
years living in the same place	16,822	41.854	40.218	45.341
Have HIV	38,514	0.082	0.092	0.049
Have HIV (Urban)	21,696	0.096	0.110	0.055
Have HIV (Rural)	16,818	0.063	0.070	0.042
Have HIV (Age<= 24)	16,859	0.044	0.054	0.015
Have HIV ($Age > 24$)	21,655	0.111	0.122	0.076
Have HIV (higher education)	15,738	0.057	0.057	0.035
Have HIV (low-Educated)	18,390	0.104	0.109	0.081
Have HIV (not-Educated)	4,417	0.082	0.091	0.028
HIV knowledge and sexual behaviour : DHS standard data				
Reduce HIV by Condom use	97,293	0.727	0.705	0.778
Reduce HIV by having one sexual partner	97,293	0.809	0.791	0.852
Healthy people can have HIV	97,293	0.787	0.768	0.834
HIV transmited by mosqito	97,293	0.774	0.761	0.803
HIV transmited by food	97,293	0.693	0.684	0.714
HIV index	97,293	3.789	3.709	3.980
Modern contraceptive use	96,425	0.301	0.291	0.329
Traditional contraceptive use	96,425	0.050	0.051	0.046
Condom use	96,425	0.120	0.082	0.206
Have non-spousal partners	73,975	0.272	0.230	0.364
Involved in a polygamous marriage	46,965	0.167	0.200	0.083
Age at first intercourse	78,324	17.722	17.521	18.199
Marriage	97,289	0.599	0.635	0.513
Fertility rate	58,054	0.103783		
Internet use: Afro-barometer				
Daily internet use	12,555	0.067	0.055	0.078
Any internet use	12,555	0.166	0.140	0.192

Note: Summary statistics are provided for the largest sample used in the regressions for each data set.

Table 1-A.3: Descriptive statistics for variables used to construct the wealth index

	Mean	Standard deviation	Min	Max
Have car	0.053	0.211	0	1
Have refrigerator	0.136	0.324	0	1
Have telephone	0.045	0.195	0	1
Have radio	0.639	0.457	0	1
Have shared toilet	0.469	0.458	0	1
Have electricity	0.279	0.426	0	1
Have bicycle	0.396	0.465	0	1
Have bed net	0.623	0.406	0	1
Roof earth	0.218	0.372	0	1
Roof tile	0.247	0.389	0	1
Roof leaves	0.174	0.342	0	1
Roof cement	0.084	0.252	0	1
Roof palm	0.006	0.068	0	1
Roof Other	0.002	0.044	0	1
Wall brick	0.104	0.25	0	1
Wall mix	0.096	0.241	0	1
Wall palm	0.043	0.165	0	1
Wall wood	0.162	0.3	0	1
Wall earth	0.027	0.132	0	1
Wall Other	0.01	0.083	0	1
Wall stone	0.102	0.247	0	1
Floor earth	0.497	0.476	0	1
Floor cement	0.06	0.227	0	1
Floor dung	0.031	0.166	0	1
Floor parquet	0.025	0.149	0	1
Floor wood	0.003	0.056	0	1
Floor vinyl	0.017	0.123	0	1
Floor other	0.001	0.033	0	1
Floor palm	0.013	0.106	0	1
Cook charcoal	0.068	0.252	0	1
Cook fire	0.368	0.459	0	1
Cook gas	0.002	0.043	0	1
Cook LPG	0.036	0.179	0	1
Cook dung	0.476	0.475	0	1
Cook other	0.002	0.041	0	1
Cook karo	0.015	0.117	0	1
Cook bio	0.005	0.069	0	1
Cook coal	0.008	0.084	0	1
Cook electricity	0.012	0.104	0	1

Table 1-A.4: Effect of the number of Internet nodes on the search popularity of some sex-related topics

Search popularity of:	(1)	(2)	Number of observations
HIV/AIDS	0.0229***	0.0107***	121
	(0.00208)	(0.00265)	
Marriage	0.0266***	-0.00577*	121
	(0.000893)	(0.00340)	
Pregnancy	0.0313***	0.00632**	121
	(0.00264)	(0.00282)	
Sex	0.0223***	0.00880**	121
	(0.000850)	(0.00340)	
Condom use	0.0259***	0.00572**	121
	(0.00234)	(0.00219)	
Sexual abstinence	0.00415***	0.000716	110
	(0.000685)	(0.00309)	
Sexual abuse	0.0238***	0.00223**	110
	(0.000952)	(0.00102)	
Sexual words	0.0263***	0.00728***	121
	(0.00138)	(0.00212)	
Sexually transmitted infections	0.0171***	0.00756***	121
	(0.00144)	(0.00187)	
Polygamy	0.00455***	0.00884***	110
	(0.00100)	(0.00168)	
Sexual partner	0.0236***	0.0118**	121
	(0.00196)	(0.00594)	
Porn	0.0142***	0.00505**	121
	(0.00174)	(0.00258)	
Country FE	No	Yes	
Year FE	No	Yes	

Notes: Data are from Google Trends. Searching interest in a specific topic is an index between 0 to 100 and represents relative interest in the topic compared to the maximum volume of search activities in a specific time and location. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 1-A.5: Estimates, P values, and 95% confidence intervals for alternative standard errors

Treatment group	Timing of submarine arrival	
Estimate	-0.022	
Cluster-robust s.e. (grid 10 *10 level)	0.011	
t-statistic	-2.05	
P values and confidence intervals:	P val.	Conf. int
Bootstrap by grid 10 cells(restricted)	0.1250	[-0.0642, 0.003855]
Bootstrap by grid 10 cells(unrestricted)	0.1875	[-0.07362 , 0.02824]
Bootstrap by by grid 10 cells-country (restricted)	0.1344	[-0.05676, 0.008306]
Bootstrap by grid 10 cells-country(unrestricted)	0.1429	[-0.05827 , 0.01289]
Bootstrap by by grid 10 cells-year(restricted)	0.1427	[-0.05934, 0.01021]
Bootstrap by grid 10 cells-year(unrestricted)	0.1489	[-0.05988, 0.0145]
Bootstrap by by individual(restricted)	0.1474	[-0.05894, 0.01321]
Bootstrap by individual(unrestricted)	0.1490	[-0.05934, 0.01397]

Table 1-A.6: Countries' 2021 population and Internet users statistics

AFRICA	**Pop(2021)**	**IntUsers (2000)**	**IntUsers (2020)**	**IntPen**	**IntGrowth**	**Facebook subscribers (12/31/2020)**
Algeria	44,616,624	50,000	25,428,159	57.00%	50756%	25,140,000
Angola	33,933,610	30,000	8,980,670	26.50%	29835%	2,831,000
Benin	12,451,040	15,000	3,801,758	30.50%	25245%	1,538,500
Cameroon	27,224,265	20,000	7,878,422	28.90%	39292%	4,267,000
Congo	5,657,013	500	833,200	14.70%	166540%	833,200
Egypt	104,258,327	450,000	54,741,493	52.50%	12064%	48,830,000
Ghana	31,732,129	30,000	14,767,818	46.50%	49126%	7,944,000
Kenya	54,985,698	200,000	46,870,422	85.20%	23335%	10,444,000
Madagascar	28,427,328	30,000	2,864,000	10.10%	9446%	2,864,000
Morocco	37,344,795	100,000	25,589,581	68.50%	25489%	21,730,000
Namibia	2,587,344	30,000	1,347,418	52.10%	4391%	792,000
Nigeria	211,400,708	200,000	203,168,355	96.10%	101484%	31,860,000
Senegal	17,196,301	40,000	9,749,527	56.70%	24273%	3,802,000
South Africa	60,041,994	2,400,000	34,545,165	57.50%	1339%	24,600,000
Sudan	44,909,353	30,000	13,124,100	29.20%	43647%	1,300,000
Tanzania	61,498,437	115,000	23,142,960	37.60%	20024%	5,223,000
Togo	8,478,250	100,000	1,011,837	11.90%	912%	860,500
Uganda	47,123,531	40,000	18,502,166	39.30%	46155%	3,328,000
Zambia	18,920,651	20,000	9,870,427	52.20%	49252%	2,543,000
AFRICA	1,373,486,514	4,514,400	634,863,323	46.20%	13963%	255,412,900
Rest of World	6,502,279,070	356,471,092	4,463,600,449	68.60%	87.50%	2,474,111,201
WORLD	7,875,765,584	360,985,492	5,098,463,772	64.70%	100.00%	2,729,524,101

Notes: Africa Internet statistics for December 31, 2020 have been updated as of February 14, 2021. Africa Facebook subscribers are estimated for December 31, 2020. Africa Population numbers are mid-year 2021 estimates based on data from the United Nations Population Division, and Africa Internet usage information comes from, among others, data published by WWW, ITU , Facebook, and other trustworthy information sources. The data source for this table is Internet World Stats (www.Internetworldstats.com.)

Chapter 2

Internet Penetration and the Financial Behaviour of Adults in the U.S.

2.1 Introduction

In recent years, financial markets have become more accessible to small investors in the United States. According to the Federal Reserve, 56 percent of American families held stocks in 2021, compared to only 32 percent in 1989 (Federal-reserve (2022)). We know very little about the factors underlying this increase in financial market participation. Microeconomic models of financial decision-making generally assume that agents who make financial decisions do so rationally and in accordance with their self-interest. This framework assumes that individuals are financially literate and understand which financial actions promote their interests. Empirically, it has been found that individuals with higher levels of financial literacy are more likely to benefit from participation in the stock market (Lusardi and Mitchell (2014); Lusardi et al. (2017)). In reality, however, few individuals are financially literate. This lack of financial literacy predominantly occurs because acquiring financial knowledge comes at a cost (Vissing-Jørgensen (2002); Campbell (2006)). For example, participation in the stock market requires individuals to know how to collect information on stock market opportunities (Guiso and Jappelli (2005)), set up a trading account as an investor, and learn how to do trading (Barber and Odean (2001)). It follows that reducing the cost of access to financial information is likely to foster participation in financial markets. This hypothesis, however, has seldom been tested. In this paper, we test it by investigating the causal impact of increased access to fast and cheap Internet on financial behaviours.

We leverage information from a recent program aimed at increasing access to high-speed Internet in the United States. In 2010, the Federal Communications Commission (FCC) released the National Broadband Plan (NBP), which aimed to provide every American with "affordable access to robust broadband service" by 2020 FCC (2019). The NBP adopted a new, expanded Universal Service Fund, embracing affordable broadband with at least 4 Mbps of actual download speed and 1 Mbps of actual upload speed. The subsequent broadband expansion revolutionized access to information and financial services, including electronic banking, financial planning, and financial forecasting tools. Online investors conduct many brokerage activities directly and more cheaply (Fernández-López et al. (2018)); until 2021, around 65 percent of Americans did online banking, and 38 percent had a financial advisor.

By reducing the cost of acquiring information, we hypothesize that access to fast and cheap financial services via high-speed broadband Internet will improve financial literacy and increase participation in financial markets. Numerous studies have investigated the link between financial literacy and financial behaviour (Lusardi and Mitchell (2007); Skimmyhorn

(2016)). Experimental studies have also documented the relationship between financial literacy and stock market participation (Drexler et al. (2014); Bruhn et al. (2016); Carpena et al. (2019)).[1] However, no study has investigated the causal impact of access to the Internet as a main source of information on financial literacy. Further, despite the increasing importance of the Internet in the United States and other countries around the world, studies that analyze the implications of this technology for households' financial decisions from a supply-side perspective are still preliminary. Earlier studies on the effect of the Internet on financial behaviour have mainly focused on the effects of Internet use. Also, the literature analyzing the impact of the Internet on financial decisions has primarily focused on stock market participation (see, e.g., Bogan (2008); Liang and Guo (2015); Fernández-López et al. (2018); Hvide et al. (2020)). The literature has yet to explore the effects of the Internet on other important financial activities, such as having a mortgage or having an active checking or savings account. Moreover, possible interconnections between financial decisions and how they relate to improvement in information and communication technologies have not been sufficiently analyzed.[2] Finally, one of the critical caveats related to previous studies is that their data does not allow to use of identification strategies which account for unobserved individual-level heterogeneities in Internet access and financial decision-making. This failure to account for the endogenous nature of Internet access is likely to result in biased estimates.[3]

In this study, we fill these gaps in the literature by estimating the causal impact of broadband Internet expansion on individuals' participation in the stock market, the probability of having a mortgage, and the probability of having an active checking or saving account. To understand the possible mechanisms behind these effects, we also analyze the effect of broadband expansion on individuals' financial literacy and their use of online and offline financial services such as online banking or having a financial advisor. Additionally, we document significant heterogeneity in the effect of broadband expansion on stock market participation. We analyze heterogeneous effects by gender, age, marital status, education, income, level of social connection, and mental health status. Furthermore, we examine the interconnections between different financial decisions to find how broadband diffusion affects stock market participation with the possibility of investment in housing or savings in the bank. Our analysis provides significant insights into how the Internet has affected financial inclusion in the

[1] Fernandes et al. (2014) and Kaiser and Menkhoff (2017) summarize the empirical literature in this field.

[2] For example, following the expansion of access to the mortgage due to technological improvement and the resulting high rate of mortgage use, the exit rate from the stock market has significantly increased in the U.S. (Fuster et al. (2019)). Around 20% of households in the U.S. with an outstanding mortgage owned no stocks during the two years after purchasing their first home (Brandsaas (2021)). This suggests that the high rate of mortgage used in the U.S. affects other financial decisions, such as stock market participation.

[3] A notable exception is Hvide et al. (2020), who exploit plausibly exogenous variations in Broadband use provided by a public program in Norway to analyze the effect of this technology on stock market investment.

United States.

To investigate the impact of access to high-speed Internet on financial behaviour and possible mechanisms behind this effect, we merge individual-level longitudinal data on financial activities, demographic information, and health status obtained from the Health and Retirement Study (HRS) for the 2010 to 2016 time period [4], with the National Broadband Map on residential high-speed broadband penetration provided by the National Telecommunications and Information Administration (NTIA) in conjunction with the Federal Communications Commission (FCC). [5] Since the Health and Retirement Study surveys the same individuals over time, it can be used to analyze how an individual's financial decisions respond to changes in Internet penetration over time. Our sample follows 9,561 individuals during the 2010 to 2016 survey years.

Because the diffusion of broadband infrastructure may not have been random, we use two identification strategies to address this potential endogeneity issue. The first strategy, the baseline strategy, exploits both within-county and within-individual variations in broadband diffusion. For the baseline strategy, we estimate an OLS regression that controls for individual, county, and survey-year fixed effects. Controlling for individual fixed effects is possible because of the longitudinal nature of our dataset. Accounting for individual fixed effects controls for all individual-level characteristics which might affect Internet access and financial decisions that do not change over time. We also control for a wide range of time-variant individual and household-level characteristics. Our second identification strategy uses an instrumental variable (IV) approach to account for the aforementioned endogeneity issue. In the IV model, we use lightning activity (cloud-to-earth strikes) as an instrument for Internet penetration. Using National Aeronautics and Space Administration (NASA) satellite-generated data on the incidence of lightning for the entire U.S., we construct lightning density data per square kilometre at the county level with a one-year lag from 2009 to 2015. This instrument can generate functional variations in broadband Internet penetration for the following year. The rationale underlying our choice of instrument is that Internet penetration is highly sensitive to power disruptions (Agrawal (2021); Manacorda and Tesei (2020); Andersen and Dalgaard (2013); Andersen et al. (2011)). Frequent electrostatic discharges during storms damage Internet infrastructure and negatively affect Internet connectivity (Andersen et al.

[4] The year 2010 is the year in which the NBP plan was launched.

[5] The FCC statistical reports on broadband deployment are summarized data about high-speed broadband Internet subscribership that delivers at least 10 Mbps downstream and at least 1 Mbps upstream per 1000 Households, which was part of 2010 National broadband plan. To check the validity of our results, we also used the data on the number of Residential Fixed High-Speed Connections of over 200 kbps per 1000 Households. The high-speed residential connections are coded between 0 to 5 based on the number of households with fixed-line connections in each county.

(2012)). Power surges and falls lead to equipment failure and damage, and the frequency of power failures increases maintenance costs, reducing the incentive for Internet expansion. In the first stage, we find strong evidence that the frequency of lightning strikes negatively affects broadband penetration. Our first-stage estimates suggest that a one-unit increase in the average density of lightning strikes per square kilometre in the previous year decreases Internet penetration in the current year by 0.0218 index points. This analysis controls for individual, county, and survey-year fixed effects and all the time-variant variables controlled for in the baseline strategy.

Both identification strategies show a significant effect of broadband Internet on financial decision-making. The OLS estimates show that an increase in broadband penetration by one index point (out of five) increases the likelihood of stock market participation by 3 percentage points, the probability of having a mortgage by 3.1 percentage points, and the probability of having active savings account by 4.8 percentage points. The analysis of the interconnection between financial decisions reveals that the probability of investment in the stock market is lower for an individual who has a mortgage payment and does not have an active savings account.

The instrumental variable estimates are generally consistent with our OLS estimates with higher magnitude. OLS estimates provide the average treatment effect (ATE) over the entire population, while the IV estimates provide the local treatment effect (LATE). The IV specification estimates the behaviour of a subgroup of individuals for whom the investment in stock markets and housing or having active saving habits is greater than average (Ciacci (2021); Bhuller et al. (2020); Angrist and Imbens (1995)). The IV estimates show that an increase in broadband penetration by one index point (out of five) increases the likelihood of stock market participation by 4.3 percentage points, the probability of having a mortgage by 4.3 percentage points, and the probability of having active savings account by 6.2 percentage points. These estimates are larger than the OLS estimates, but the differences are not big.

To address concerns about the validity of the identification assumption used in our IV specification, we performed a placebo test analyzing the effect of broadband penetration in 2016 on the level of stock market participation in 1992, when broadband Internet was unavailable. We found no correlation between broadband penetration in 2016 and the level of stock market participation in 1992. The results of this test suggest that stock market participation has no causal effect on broadband expansion. Additionally, to check if lightning density affects stock market participation through channels other than access to broadband Internet, We perform an additional check for correlations between lightning density and other potential determinants of stock market participation, such as unemployment and local economic growth. We find no statistically significant relationship between lightning strike density and unemployment and a positive and significant correlation between lightning density and economic growth at the county level. The significant correlation between flash density and GDP growth is not the source of concern. If a higher lightning strike increases production and economic growth, we can expect a positive relationship between lightning density and stock market participation. Our results do not confirm this effect. Hence, the instrumental variable satisfies the exclusion restriction test, which is necessary for choosing IV. Further, following Conley (1999), we account for spatial correlation between standard errors within 100 meters for our main variable of interest, which is stock market participation. The result represents spatial correlation among standard errors has not biased the results.

Next, we analyze possible heterogeneity in the effect of broadband penetration on stock market participation. We find that this effect is stronger for men, younger individuals, people who are married, are not retired, are not depressed, have a higher level of education, have a higher income, and individuals who are more socially connected. The heterogeneous analysis suggests that the effect of broadband access on stock market participation is stronger for individuals with more social and financial privileges.

Finally, We test several possible mechanisms through which broadband penetration might affect financial decisions. First, we find that higher broadband penetration increases daily Internet use; increasing broadband diffusion by one index point increases Internet use by 4 percentage points. Second, we find that broadband penetration increases financial literacy. In order to estimate the effects of the Internet on financial literacy, we construct an index score using five questions from HRS that assess respondents' general financial knowledge. A respondent earns one point for each correct answer, so theoretically, the index takes values between 0 and 5. We find that an increase in broadband penetration by one index point increases the comprehensive knowledge of finance by 1.618 index points. Third, we find that broadband diffusion increases the use of financial services, such as online banking and having a financial advisor. Specifically, we find that an increase in broadband penetration by one index point increases the use of online banking and the probability of having a financial advisor by 4.3 and 2.4 percentage points, respectively.

The rest of this paper is organized as follows. Section 2.2 presents our main contributions to the literature. Section 2.3 presents the conceptual framework, which describes the theoretical effects of broadband penetration on financial behaviours. Section 2.4 describes the context of the study, the data, and the descriptive statistics. Section 2.5 presents our empirical strategies. Section 2.6 presents our results. Section 2.7 discusses a placebo test and shows evidence supporting our identification assumptions. Section 2.8 concludes the paper.

2.2 Contribution to the literature

Our paper documents the causal effect of access to broadband Internet on financial decisions and the interconnections between these decisions in the United States. It is also the first to demonstrate significant heterogeneity in this effect by mental health status, gender, age, education, marital status, income, and sociability. The paper also identifies several mechanisms that might contribute to these underlying effects. Although we analyze a more extensive set of outcomes, we view our study as contributing to the body of literature that documents the drivers of stock market participation. An early empirical study in this literature is the work of Hong et al. (2004). They found that having a larger number of social connections increases an individual's probability of stock market participation. Subsequent studies have also found that other factors such as IQ (Grinblatt et al. (2011)), wealth and income (Brunnermeier and Nagel (2008); Calvet et al. (2009); Briggs et al. (2021); Fagereng et al. (2021)), financial literacy (Lusardi and Mitchell (2014)), educational attainment (Cole et al. (2014); Lusardi et al. (2017)), and computer ownership (Bogan (2008)), also affect an individual's probability of participating in the stock market. Our work contributes to this literature by showing that availability of broadband Internet increases investment in the stock market by reducing the fixed costs of market participation through access to information and financial services. However, as already noted, we differ from previous studies by documenting significant heterogeneity in this effect.

We specifically contribute to the literature that documents the impact of the Internet on stock market participation. An early study in this literature is Bogan (2008). This study found evidence suggesting that greater access to the Internet reduces financial friction and increases stock market participation. However, later studies by Glaser and Klos (2013) and Leung (2013) found that Internet use did not have a significant effect on stock market participation and that the observed increase in stock market participation occurred because of changes in other factors such as experience and knowledge, unrelated to Internet expansion. More recent studies by Fernández-López et al. (2018) and Liang and Guo (2015) examine whether the use of the Internet is a complement or substitute for the effect of social interactions on stock market participation. Liang and Guo (2015) consider in-person social interactions, whereas Fernández-López et al. (2018) consider online social networks. Both studies find that using the Internet reduces social interaction's effect on stock market participation. Using data from Norway, Hvide et al. (2020) find that Internet use induces individual investors to make better financial decisions, also leading to a "democratization of finance." Our work differs from previous studies in the identification strategies and principal conclusion. Using longitudinal data enables us to control for individual fixed effects, and we choose a plausibly exogenous instrument which was not done before. While we find that broadband diffusion increases participation in the stock market, our heterogeneity analysis shows that this effect is more prominent among certain individuals.

Our paper also contributes to the literature examining the impact of financial literacy on other financial decisions such as saving, portfolio choice, investment and retirement planning (Agarwal et al. (2011); Agnew and Szykman (2011); Van Rooij et al. (2011); Lusardi and Mitchell (2008); Luhrmann et al. (2018)). These studies have mainly considered education as the primary determinant of financial literacy (Bernheim et al. (2001); Hilgert et al. (2003)). Unlike these studies, we investigate the role of Internet access, rather than education, as a potential determinant of financial literacy.

Finally, our finding of a positive relationship between broadband penetration and having a financial advisor relates our study to the literature investigating the determinants of financial advisor use. Previous studies in the literature have established that age (Agarwal et al. (2009)), income and wealth levels (Grable and Joo (2001); Hanna (2011); Bluethgen et al. (2008); Chang (2005)),gender (Bluethgen et al. (2008); Grable and Joo (2001)), individual personality (Kramer (2016); Xiao and Porto (2016)), and psychological factors such as anxiety (Gerrans and Hershey (2017)) and trust (Burke and Hung (2021); Stolper (2018)) are important determinants of financial advisor use.

2.3 Conceptual framework

In this section, we investigate the theoretical channels through which broadband Internet access can affect financial decisions, such as investment in stocks, ownership of mortgages, and individuals' propensity to save. Figure 2-4 summarizes these possible channels. With lower associated prices and higher speed, broadband Internet has reduced financial market frictions, such as the cost of accessing financial information and the fixed costs of financial market participation. Broadband availability has increased the home Internet take-up rate (De Stefano et al. (2014); Chinn and Fairlie (2010)) and made financial markets more accessible to small investors. In 1998, only 18 percent of households in the U.S. had home Internet, compared to around 89 percent in 2019. Regular Internet users have access to a plethora of online financial information. For example, online learning has provided Internet users with a practical method for acquiring financial information (Wolla (2017); Cao and Liu (2017)). Additionally, access to social media and other online social networks has allowed individuals to easily exchange financial information online (Fernández-López et al. (2018)). This broad access to financial information will likely increase financial literacy considerably among Internet users (Sabri and Aw (2019)).

The current technological growth and de facto standardization of the Internet also contributed to an increasing array of low-cost and easily accessible online financial services. For example, online banking has benefited both account holders and financial institutions. As such, the popularity of online banking has increased considerably in the United States, with approximately 64.6 percent of U.S. bank customers using online banking services as of 2021 (Statista (2022)). Previous studies indicate that households using online banking are, on average, better managers of their household finances (Hogarth and Anguelov (2004)). Additionally, access to digital financial services via mobile phones has considerably increased financial inclusion. It has been argued that the high cost of accounts and the extended distance from financial institutions are the main barriers to the use of financial services for individuals residing in rural areas. Therefore, access to the Internet, and thus online financial services, is likely to significantly increase access to financial services for individuals in remote areas (Andrianaivo and Kpodar (2012); Okello Candiya Bongomin et al. (2018)).

Access to the Internet also delivers other financial services, such as online access to financial advisors. Similarly to the availability of online banking, online access to financial advisors is likely to disproportionately benefit individuals in remote areas who, due to physical distance, lack access to these services. Additionally, online financial advisors give clients access to additional financial information which is unavailable to the general public (Wisniowski and Lindstrom (2012)).

In 1998 about 21% of households in the U.S. used a financial advisor, and by 2007 this figure had increased to 25% (Hanna (2011)). According to a survey of 2300 Americans conducted in 2020 by the Harris Poll on behalf of Northwestern Mutual, about 30% of individuals consult with financial advisors. Most individuals who hire financial advisors are older adults with considerable wealth but may lack the ability to manage their resources properly (Agarwal et al. (2009)). However, numerous factors other than access to the Internet, including income (Hanna (2011); Grable and Joo (2001)), education (Hanna (2011)), and Wealth (Chang (2005); Bluethgen et al. (2008)) also impact the use of financial advisors (Cummings and James III (2014)). Additionally, females are more likely to use financial advisors than their male counterparts (Grable and Joo (2001); Bluethgen et al. (2008)). Therefore, we expect the Internet's effect on financial decision-making to vary across demographic and socioeconomic characteristics.

Internet access is also likely to affect participation in the stock market, mortgage ownership, and the interaction between these two financial activities. Indeed, after the emergence of high-speed Internet and online trading in the 1990s, many major U.S. financial firms developed sizeable online financial services such as providing online information on stock prices and online financial analysis tools (Bogan (2008)). Previous studies have found that the provision of these online services has had a significant impact on the stock market participation rate (Hong et al. (2004); Bogan (2008); Hvide et al. (2020)). Beyond stock markets, increasing Internet access has led to technological innovation in the residential mortgage industry. After the 2008 financial crisis, online FinTech lenders, for example, have become an increasingly essential source of mortgage credit to U.S. households. [6] As of December 2016, all American FinTech lenders were non-bank lenders that scrutinized mortgages and worked without deposit financing or branch networks. Their lending has grown annually by 304%, from 34 billion dollars of total origination in 2010 (2% of the market) to 161 billion dollars in 2016. Their growth has been particularly pronounced in mortgage refinancing, insured by the Federal Housing Administration (FHA) (Fuster et al. (2019)). As of 2021, over 65 percent of U.S. residents were homeowners with outstanding mortgages. Online access to the housing market increased competition among start-ups and existing lenders, reducing interest rates meant as firms competed to attract new customers. The Internet also reduced mortgage lending frictions and loan processing times, increased capacity constraints, and access to finance for some borrowers. Lenders embraced digital mortgages more than before. In 2022 around 45% of Americans used the Internet in the entire process of applying for a mortgage to buy a property, 10% used paperwork for most of the process, and 0.3 did the whole application process on paper (Richardson (2022)). Following these changes, we expect a positive relationship between Internet penetration and investment in housing.

As already noted, the Internet may also affect differently on financial decisions depending on how much money they can allocate to a specific financial decision. In other words, the interconnection between different financial decisions is a crucial aspect that has been largely overlooked in the literature examining the determinants of financial decisions. In the United States, few individuals have the wealth required to purchase a home outright, and as a result, over 62 percent of Americans carry a mortgage. In 2020, American Homeowners owed 10.3 trillion USD in mortgages, and on average, 208,185 USD per mortgage.[7] Theoretically, we expect individuals with an outstanding mortgage to participate less in the stock market than the rest of the population. This is because individuals with an outstanding mortgage are required to make monthly mortgage payments and, as such, may have less disposable income, reducing their ability to participate in the stock market and their incentive to do so. Indeed, previous studies indicate that delaying home ownership has a positive and statistically significant effect on individuals' level of investment in the stock market (Fagereng et al.

[6] FinTech lenders offer an application process that can be completed entirely online.

[7] U.S. Census Bureau 2020.

(2017); Vestman (2019)). Also, renters' stock market participation rate is likely to be higher than that of homeowners as they have not put a significant portion of their personal wealth into down payments. In line with these observations, we expect the impact of Internet access on stock market participation to be more muted for individuals with outstanding mortgages.

2.4 Context, data, and descriptive statistics

2.4.1 Broadband penetration in the US

The last three decades have witnessed a dramatic expansion in the availability of the Internet. Between the introduction of the commercially available Internet in the mid-1990s and the turn of the 21st century, most residential access to the Internet in the U.S. was via dial-up connections. However, dial-up speeds were too slow to download large files or stream sound clips or videos, and access meant tying up telephone lines. By the late 1990s, the fastest dial-up connection available to most users was only 56 Kbps. Broadband began to diffuse after the 1996 telecommunications act, and the use of dial-up connections subsequently started to decline. In the early 2000s, broadband technology was developed further and became the most influential tool used for daily communications, activities, and transactions. In 1999 the FCC defined broadband as any Internet connection with a speed exceeding 200 kbps. During that period, most web pages were almost text-based, with little embedded graphics or video; as such, the 200 kbps threshold was arguably a sufficient benchmark for broadband capability at the time. Following the widespread implementation of data-intensive features, such as rich graphics and embedded videos, in most websites and the emergence of numerous websites that were primarily created to serve video content, the 200 kbps threshold was no longer serving as a meaningful definition of broadband Internet.

On March 16, 2010, the Federal Communications Commission (FCC) released "Connecting America: The National Broadband Plan (NBP)", mandated by the American Recovery and Reinvestment Act of 2009 (ARRA, P.L. 111-5). The FCC's National Broadband Plan was a 360-page document composed of 17 chapters containing 208 specific recommendations directed to the FCC, the Executive Branch (both to individual agencies and the Administration as a whole), the Congress, and non-federal and non-governmental entities. The ARRA specified that the NBP should "seek to ensure that all people of the United States have access to broadband capability" by removing barriers to infrastructure investment and promoting competition in the telecommunications market until 2020. The NBP adopted a new expanded Universal Service Fund definition embracing affordable broadband with at least 4 Mbps of actual download speed and 1 Mbps of actual upload speed (Wohlman (2020)).

Following the national plan, access to broadband Internet started to grow more quickly. Many providers began to expand high-speed Internet infrastructures. Figure 2-6 illustrates the number of total and residential fixed Internet access connections over 200 kbps in at least one direction (excluding mobile wireless connections). The figure also shows the total number of fixed connections per 100 persons and the number of residential fixed connections per 100 households since 2001. In 2010, the total number of fixed connections (business and residential) was 84 million connections and reached 106 million connections By December 2019. In 2009, only 5 percent of all U.S. households were using dial-up, while 65 percent had access to Internet connections with speeds exceeding 200kbps in at least one direction. The subscription rate among households reached 82 percent in 2016. In 2019, 93.5 percent of Americans had access to high-speed broadband Internet. FCC (2019)

Figures 2-1 and 2-2 show the maps for the number of residential fixed Internet connections exceeding 200 kbps in at least one direction per 1,000 households for the years 2010 and 2016. In 2010, the average score of the broadband penetration index was 3.8 (out of 5), meaning that, on average, between 400 to 600 of American families in each county had access to broadband Internet services exceeding 200 kbps. In 2016 the index reached 4.35, representing 600 to 800 families having access to high-speed Internet.

2.4.2 Internet use and financial activities in the U.S.

Following the expansion of high-speed broadband infrastructure, the broadband adoption rate increased over time. Internet use in the United States increased rapidly between 2000 and 2020. In 2000, about half of all American adults were using the Internet. By 2020, Internet use among US adults had increased to around 93%. Similarly, the fraction of American adults with high-speed broadband access increased rapidly. However, the growth in broadband access has slowed in recent years, as most Americans already have access to broadband services. In 2020, roughly three-quarters of American adults had broadband Internet service at home and the adoption varied across age groups. Figure 2-5 represents the rate of broadband Internet adoption and Internet use across different age groups in the U.S. Interestingly, for individuals over 50 years old, broadband adoption rates reached 71 percent in 2021, while for those 18-29 years old, it had reached 70 percent. The most rapid growth in Internet use is among adults over the age of 65. A study conducted by Fisher et al. (2016) on U.S. adults over the age of 55, using data from the Health and Retirement Study, shows that recent generations of elderly adults are more familiar with the technology. They found that the rate of Internet use for individuals over the age of 55 increased from 32% in 2002 to 61 % in 2014.

Stock market participation of Americans is increasing during the time but it is acknowledged that participation rates are much lower than what was predicted by traditional financial models, a fact which is often referred to as the "stock market participation puzzle." Between 2001 and 2008, the stock market participation rate in the United States was fairly high, as roughly 62 percent of U.S. adults owned stock during this time period. However, the stock market participation rate dropped significantly following the 2008 financial crisis and not has yet recovered to pre-2008 levels. As of 2021, around 56 percent of Americans reported holding stock, mainly through retirement accounts such as 401(k)s.

Stock ownership rates are strongly correlated with household income, education level, age and race. A 2021 study by the Federal Reserve (FED) found that only about one-third of families in the lower half of the income scale had stock holdings (Saad (2019)). This study also found that about 70 percent of families with household incomes between the 50th and 90th percentile owned stock, while families with incomes above the 90th percentile had stock ownership rates above 90 percent. Participation in the stock market varies considerably across income groups, but even among those with annual family incomes of less than $35,000, about one in five have assets in the stock market. This share increases as income rises. Among US Households with incomes above $100,000, 88 percent own stocks, either directly or indirectly. Families headed by non-Hispanic white adults are more likely than those headed by Hispanic adults or non-Hispanic black adults to invest in the stock market. Around 61 percent of non-Hispanic white households owned stock, this is relative to just 31 percent of non-Hispanic black households and 28 percent of Hispanic households. Rates of stock market participation vary by age. Among families headed by a young adult (those under 35), 41 percent own some stock, either directly or indirectly. This is true of a majority of households headed by those ages 35 to 64.

Investment in real estate or housing is an incredibly important financial decision. For many Americans, the dream of owning a house can seem far-fetched. In the past, it was possible for families to save enough money to buy properties outright. Due to rising property prices and years of economic uncertainty, it is only possible with the help of a mortgage. Hence, mortgage debt is the primary way of financing home ownership in the U.S. In the United States, more than 65 percent of adults are homeowners with outstanding mortgages. Among these people, about 19 percent are above 56 years old.[8] According to the Federal Reserve data, mortgage debt has grown over time. In 1945 total mortgage payment was 35,853 million dollars. It gradually increased and reached its highest point of 14.7 trillion dollars in 2008. Despite a short decrease after the bursting of the U.S. housing bubble and the subsequent global financial crisis. After recovering from the financial crisis, the total amount

[8]U.S. census 2021.

of mortgage debt has been growing until recent years. In 2021, the total US mortgage debt amounted to 18 trillion USD.[9] Before the introduction of the Internet, newspaper ads were the primary information source for real estate sales. However, the widespread adoption of the Internet has caused the real estate industry to change dramatically (Swanepoel (1998)). The U.S. residential mortgage industry is experiencing a wave of technological innovations after the expansion of the Internet with affordable prices and speed, as borrowers take advantage of low-interest rates; it increases the competition among start-ups and existing lenders to offer lower interest rates to attract new customers. Moreover, it reduces frictions in mortgage lending, such as lengthy loan processing, capacity constraints, inefficient refinancing, and limited access to finance by some borrowers. In 2020, the Internet was the primary source of information consumers used when buying a house. In that year, about 97 percent of homebuyers used the Internet in their home search. Meanwhile, the mortgage industry has been gradually adopting digital technology to streamline the front-to-back process of getting a mortgage. The rise of digital technology ushered in a new era for the mortgage application process, as borrowers took advantage of historically low-interest rates and lenders embraced digital mortgages more than ever before. Figure 2-7(d) shows the positive correlation between Internet penetration and the probability of owning a mortgage in our data set. The graphs indicate that the strength of this correlation increased between 2010 and 2016.

Personal savings as a percentage of disposable personal income (DPI), frequently referred to as "the personal saving rate," is calculated as the ratio of personal savings to DPI. In the short term, money is usually put aside for upcoming expenses or purchases. In the long term, the main reason behind saving might be buying a house, saving for college, or saving for retirement. Some households decide to invest saved money in various financial instruments such as mutual funds, savings bonds or equities. Between the 1970s and to 2008 financial crisis, the personal saving rate in the United States decreased from about 10%-13% to 3%-4%. The savings rate subsequently rebounded Between the end of the financial crisis and the beginning of the COVID-19 Pandemic, levelling off at about 7%. Savings spiked at the onset of the COVID-19 pandemic, increasing rapidly to 16% in 2020, as consumer spending plummeted. The personal saving rate decreased to 5.4, the rate of pre-pandemic levels. The vast diffusion of the Internet with affordable prices during recent years provided lots of financial services and helped individuals to find the highest interest rate for savings and high accessibility to their accounts. Financial literacy affects the household saving rate (Georgarakos et al. (2015)). Digital finance makes individuals more likely to start first-time savings accounts and increase their savings (Saad (2019); Ky et al. (2018)).

[9] Statista Research Department, Jun 30, 2022

In 2022, most Americans (78%) preferred doing their banking digitally via a mobile banking app or bank websites, and 25 percent did savings through online banking only.[10]. Figure 2-7(c) confirms the positive correlation between Internet penetration and the probability of having an active savings account in our data set between the years of 2010 and 2016.

2.4.3 Data sources, main variables, and descriptive statistics

In order to perform our analysis, we combine longitudinal data on individuals' financial behavior, individual-level HRS data on demographic and health information, and residential broadband penetration data obtained from the National Broadband Map. Additionally, to construct the instrumental variable used in our 2SLS regressions, we use satellite data on the incidence of lightning strikes obtained from the National Aeronautics and Space Administration (NASA). We provide more information on each of these data sources and how they are used below.

Financial decisions, financial literacy, and financial service use

We use the Health and Retirement Survey (HRS) to construct our main dependent and other variables. The HRS is a nationally representative longitudinal data set collected by the Institute of Social Research at the University of Michigan. The first survey round of the HRS was conducted in 1992 and focused on collecting demographic and health information on individuals born between 1931 and 1941 (a group aged between 51 to 61 at that time). HRS cohort members have been interviewed every two years since 1992. During the years, the HRS added different cohorts to the data set to keep the balance among respondents based on the number of observations and age.[11]

The empirical analysis in this study uses confidential individual-level data on financial behaviours, taken from the 2010 to 2016 waves of the HRS. One of our primary dependent variables is individual stock-market participation. Specifically, we classify an individual as participating in the stock market if they meet the following two conditions: 1) the individual owns stocks, and 2) these stocks are held outside of IRA or Keogh accounts and are not part

[10] Forbes study 2022

[11] In 1993, the HRS added a cohort to the study, referred to as the Asset and Health Dynamics Among the Oldest Old (or AHEAD for short); this group consisted of those born before 1924 and were 70 years of age or older in 1993. In 1998, the HRS and AHEAD cohorts were merged, and two new cohorts were enrolled and added to the study; these cohorts were respectively a sample of individuals born between the years 1924 and 1930, referred to as the children of the Depression (CODA), and a sample of individuals born between 1942 and 1947, referred to as the War Babies. In 2004, the HRS added the Early Baby Boomers, a sample of individuals born between the years 1948 to 1953; in 2010, a sample of the Mid Baby Boomers, born between 1954 and 1959, was added. In 2016 the HRS added the Late Baby Boomers, a sample of individuals between 1960 and 1965.

of a 401K or a similarly defined contribution pension plan.[12] We define a dummy variable equal to one if the respondent answered yes to this question and equal to zero otherwise.

Our study also considers another dependent variable related to financial behaviour: the use of a mortgage in buying a house. Specifically, We define a dummy variable equal to one if an individual has a mortgage payment in buying their first, second or third house and zero if they do not have any mortgage loan. [13]

The last primary dependent variable of interest is a binary indicator for whether an individual has an active savings account. We define a dummy variable equal to one if an individual actively uses saving or checking accounts and equal to 0 otherwise.[14].

In order to identify the mechanisms behind the effect of broadband penetration on financial decisions, we first explore if Internet use changes following broadband diffusion. From the 2002 HRS survey, which asks respondents about their Internet use, we define a dummy variable with a value equal to one if a respondent reported using the Internet on a regular basis.[15]

Other variables we consider as a mechanism underlying the effect of broadband penetration on financial behaviours include financial literacy, and the use of financial services such as online banking or having a financial advisor. We used HRS cross-sectional data to construct these variables for specific waves that these questions have been asked of participants in the survey. The 2010 round of the HRS survey asked several financial questions in order to investigate the participants' financial literacy level. We considered five of these financial questions and defined dummy variables for each of the following equations.:
1-If the interest rate falls, will bond prices (rise/fall)?
2- Is it better to pick individual company stocks that will have better than average returns?
3- Suppose you had $100 in a savings account, and the interest rate was 2% per year. After five years, how much do you think you would have in the account if you left the money to grow- more than $102, exactly $102, or less than $102?

[12]For the module on Stock Market Participation, the HRS asks the following repetitive question in each survey: "For the next few questions, please exclude any assets held in the form of IRA and Keogh accounts. Do you or your (husband/wife/partner) have any shares of stock in publicly held corporations, mutual funds, or investment trusts?".

[13]The HRS asks the following question: "Do you or your husband, wife, or partner have a mortgage, land contract, second mortgage, or any other loan that uses this property as collateral? Please do not include home equity lines of credit".

[14]This dummy variable is defined based on the repetitive question in each wave, which is: "Do you or your husband or wife partner have any checking or savings accounts or money market funds?"

[15]We use the following question to define the variable of Internet use: "Do you regularly use the Internet (or the World Wide Web) for sending and receiving email or for any other purpose, such as making purchases, searching for information, or making travel reservations?".

4- Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After one year, would you be able to buy more than today, the same as today, or less than today with the money in this account?
5- Buying a single company stock usually provides a safer return than a stock mutual fund?
Each of these dummy variables was assigned a value of one if a respondent answered the corresponding question correctly.

We also define a comprehensive financial literacy index that, for each respondent, sums the points obtained from correct answers to these questions. Therefore, an individual's financial literacy index score ranges from zero (if no questions were answered correctly) to 5 (if all questions were answered correctly).

Trust in a financial advisor and online banking are two critical variables that we consider as possible mechanisms underlying the effect of broadband penetration on financial decisions. In 2016, the HRS asked whether individuals can trust financial advisors in their financial decisions or not. We define a dummy variable equal to one if a respondent trusts financial advisors and zero if they do not. Also, in 2014, the HRS asked respondents if they used online banking. We define a dummy variable equal to one if a respondent answered "yes" to this question and zero otherwise.

Finally, to control for individual heterogeneity in our analysis, we combined the original core HRS data with HRS RAND data, a research-friendly subset of the Health and Retirement Study, which includes cleaner versions of the most commonly used HRS variables. The RAND Corporation has rich and detailed panel data on HRS respondents' health, behaviour, socio-demographic characteristics, economic status, level of educational attainment, and labour market history. This allows us to include control variables for a range of individual and household-level characteristics in the multivariate analysis. The first set of control variables we include consists of the respondent's demographic and personal characteristics, including age, gender, race, education, sociability, marital status, working or retirement status, and a dummy variable indicating whether an individual is depressed.[16] Theory suggests that there should be time effects on portfolio choice if households perceive changes over time in the risks or expected excess returns of risky assets. Also, age likely has an influence on portfolio choice, as older investors have shorter horizons than younger investors and investment opportunities are time-varying or if older investors have less human wealth relative to financial wealth than younger investors (Bodie et al. (1992); Campbell et al. (2002)). Thus, it seems hard to rule out

[16] For demographic and personal characteristics, we may have two responses per household, one for a man and one for a woman. In such cases, we consider the information regarding the age, education, employment, and retirement status of the older household member. But for 98 percent of observations in our sample, we have only one person per household.

either time or age effects in studying portfolio choice. Cohort effects are more problematic. In principle, cohort effects could be caused by different labour market experiences that affect the ratio of cohort members to the financial wealth held by that cohort. However, this effect is unlikely to be strong in modern U.S. conditions. Cohort effects could also arise from differences in preferences across cohorts. The data cannot identify such effects without modelling them. To deal with potential age, period (time), and cohort effects, we follow Curcuru et al. (2004) and Campbell (2006) by setting one of the three effects as a random effect while treating the other two as fixed effects in a multilevel model (Yang et al. (2004)).

The second group of the control variables we include is related to household characteristics such as income and wealth[17], migration status [18], a dummy variable indicating whether they have children who are living with them. The third group of controls consists of community-level (county-level) controls such as the mean stock market participation, mortgage ownership, and use of savings or checking accounts.

Broadband penetration

To measure the evolution of U.S. broadband access over time, we obtained data from the National Broadband Map. The National Broadband Map is provided by the National Telecommunications and Information Administration (NTIA) in conjunction with the Federal Communications Commission (FCC). The FCC has detailed data on broadband subscriptions from its semi-annual Form 477 collection. The report contains data for Residential Fixed High-Speed Connections that deliver speeds of at least 10 Mbps downstream and at least 1 Mbps upstream per 1000 Households, which was part of the 2010 National broadband plan. To check the validity of our results, we also used the data on the number of Residential Fixed High-Speed Connections of over 200 kbps per 1000 Households in each county from the same report (see table 2-A.2). [19] The counties are assigned an index value ranging from 0 to 5, with a higher score indicating a higher level of broadband adoption in that country. Specifically, the index measures the number of high-speed residential connections per 1000 households in each county. An index value of 1 is assigned to counties with 1 to 200 high-speed connections per 1000 households. Similarly, counties with 200-400, 400 to 600, 600 to 800, and over 800

[17]Total wealth is the sum of all kinds of wealth held by household members minus the sum of all kinds of debt held by the household (including mortgage debt).

[18]households who invest more in stock markets may move to locations with a higher level of Internet penetration during the years of study. In order to investigate this issue, we utilized information on the migration status of the respondents and defined a dummy variable equal to 1 if a respondent never changed their place of residence during the years of study and zero otherwise.

[19]Form 477 qualifying carriers required the Internet providers to report the number of fixed-line connections in each zip code twice per year. We focus on broadband adoption by households, as household connections are the most frequently used Internet access points.

connections per 1000 households are assigned index values of 2, 3, 4, and 5, respectively. The value of zero is related to the counties with no broadband connection. The mean broadband penetration score for our entire sample is 3.38 out of 5. However, the mean value of this index score was not consistent over time, with the mean score in 2010 being 2.55 (out of 5) and the mean score in 2016 being 3.61 (out of 5).

Lightning density

In order to address possible endogeneity issues related to the non-random placement of broadband infrastructure, we use lightning density as an instrument for broadband penetration. Information on lightning density is from satellite-generated data on the incidence of lightning strikes and is obtained from the National Aeronautics and Space Administration (NASA).

We consider the mean number of lightning strikes per square kilometre for counties within the contiguous United States. This data was obtained using ground-based lightning detection sensors and global satellite data covering the period between 2009 and 2015. The mean value of lightning density is 3.66 strikes per square kilometre in a year. The top 10 states in terms of lighting strike density in 2019 were: Kansas, Florida, Oklahoma, Nebraska, Louisiana, Missouri, Iowa, Texas, Delaware, and Illinois. These states have the lowest levels of broadband penetration in our data set.

Descriptive statistics

Our sample consists of 9561 individuals for each wave and 38,244 individual-year observations. Information on four waves (2010, 2012, 2014, 2016) was included in the analysis. Table 2-1 provides summary statistics for all dependent, independent, and control variables for the largest sample size in our study, which is 8358 observations.

Male respondents make up 41.5 percent of our sample, while females constitute 58.4 percent of the sample. Regarding the respondent's highest level of educational attainment, 19.7 percent of individuals reported some high school, 25.6 percent reported high school, 29 percent reported some college-level education, and around 25.2 percent reported having a college-level education or higher. In terms of race, 65.8 percent of the sample respondents are white, 24 percent are black, and 10.1 percent are members of another race.

Around 46 percent of the survey respondents are married, and 5 percent live with a partner but are not married. We consider both groups as married and the rest of the categories as unmarried. Regarding employment and labour force participation, 50 percent of the respondents were retired, while 34.7 reported having a job at the time of the survey. With regard to the respondents' level of social connection, close to 70.2 percent of respondents attended a church at least once a month, and 52 percent had a good friend. Our analysis defines an individual as sociable if they reported having a good friend or attending a church at least once a month. By this metric, 72.6 percent of the respondents in the sample are considered sociable. We define a person as depressed if she or he reported feeling sad or depressed during the two weeks prior to the survey. According to this measure, the share of depressed people in the sample is around 13.9 percent. Approximately 45 percent of the sample lived in the same place of residence during the years of the study. The average household income in our sample is around $68,982 per year, and the average level of total wealth is $354,990.

About 49 percent of respondents reported having children who lived with them at the time of the survey. Economic characteristics of the sample at the local level represent the mean unemployment rate of 7.54 percent and the mean GDP growth per capita of 4 percent.

Data on the financial behaviour of respondents shows that approximately 19.3 percent of respondents have an investment in stocks or mutual funds. In 1992 around 27 percent of the sample had stocks and mutual funds. On average, about 30 percent of respondents had 401(k) retirement plans. Around 30.1 percent have a mortgage or investment in housing, and more than 74 percent have active savings or checking accounts. In the sample, 61 percent of individuals reported that they regularly use the Internet daily to shop, send and receive emails, and do other necessary work.

As previously mentioned, we constructed a comprehensive financial knowledge index, which ranges in value from 1-5 based on respondents' correct answers to the survey's five financial questions. The mean value for this index is 3.88 out of five and the possibility of answering correctly to each financial question changes between 64 to 88 percent in the sample. These financial questions were asked in the year 2010. Roughly 76 percent of respondents reported trusting financial advisors, while 59.1 percent reported using online banking.

Finally, the mean value of lightning density, which we use as our instrumental variable for broadband penetration, is 3.66 strikes per square kilometre, and broadband penetration is 3.88 out of 5.

2.5 Identification strategies

In order to examine the impact of broadband diffusion on financial behaviour, our identification strategies exploit variation in broadband access produced by the gradual expansion of broadband in the U.S. from 2010-2016. Our study uses two identification strategies. First, we use a set of ordinary least squares (OLS) regressions to study the association between broadband access and several measures of U.S. adults' financial behaviour. These OLS regressions control for individual and household characteristics, county fixed effects, survey-year fixed effects, and individual fixed effects. Our second strategy supplements the first strategy with an instrumental variable analysis.

2.5.1 Baseline identification strategy: OLS with individual fixed effect

Our baseline OLS specification is summarized in the following regression equation:

$$Finance_{ict} = \alpha_0 + \alpha_1 Inter_{ct} + X'_{ict}\delta + \mu_c + \lambda_k + \gamma_t + \theta_i + \epsilon_{ict} \tag{2-E.1}$$

Where the dependent variable, $Finance_{ict}$, measures the financial behaviour of individual i, in county c and survey year t. $Finance_{ict}$ is defined as a binary variable. Depending on the Financial variable being analyzed is assigned a value of one if an individual reported owning stocks, reported having an outstanding mortgage, or reported having an active checking or savings account. We also consider binary variables that indicate whether an individual trust a financial advisor or has online banking and a continuous variable measuring comprehensive financial knowledge as dependent variables in the same specification to examine the mechanism of the effect of broadband penetration on financial behaviour. The variable $Inter_{ct}$ is a categorical variable that measures the level of broadband penetration in county c in survey year t. X'_{ict} is a vector of individual, household and community-level control variables. The term μ_c is county fixed effects and controls for any time-invariant county-level factors likely to affect Internet penetration and the dependent variable. The term λ_k is cohort fixed effects and controls for all unobserved differences in financial decisions across cohorts. The term γ_t is survey year fixed effects and accounts for any time-variant differences in financial activities that might correlate with individuals' financial decisions. Controlling for time-fixed effects is essential, as adults' financial behaviours may change with age. Additionally, because Internet penetration is growing over time, time-related trends in stock market participation (or the other financial variables) may be mistakenly attributed to broadband penetration. Considering survey-year fixed effects together with cohort fixed effects along with a non-linear age control will account for age, period and cohort effects in financial decisions. Finally, we control for individual fixed effects, θ_i, to account for unobserved individual-level characteristics that do not change over time.The inclusion of individual fixed effects is possible due to the panel nature of HRS data. Including county and individual fixed effects allows us to control for a large number of potential endogeneity issues. County fixed effects control all time-invariant sources of heterogeneity at the county level, while including individual fixed effects effectively controls for factors like individual preferences and sorting into specific locations. ϵ_{ict} is the error term. We cluster standard errors at the county level to allow for unrestricted correlations in the error term across observations.

Controlling for the cohort, county, year, and individual fixed effects accounts for several sources of endogeneity in estimating the causal effect of Internet penetration on financial behaviour. However, within a county, Internet penetration may be correlated with individuals' financial decisions for reasons other than the causal effect of the former on the latter, and the OLS estimation with these controls may suffer from omitted variables bias and a reverse causality problem. This issue would not be a source of bias if the expansion of broadband Internet was randomly assigned across counties. However, due to the high installation cost required to build Internet infrastructure, broadband Internet expansion is generally not random. Additionally, adopting new technologies such as broadband Internet may be endogenous to the governance decisions (Acemoglu and Robinson (2000); Parente and Prescott (1999)). Hence, Information and Communication Technologies (ICTs) will likely be introduced in locations with notable financial markets and high participation rates. These areas generally have a higher average income, a larger share of high-skilled workers, and a younger population. Given that these socioeconomic factors are correlated with individuals' financial behaviour, this issue can result in endogeneity bias in the OLS results.

2.5.2 Instrumental variable approach

In order to strengthen our claim of causality, we use an instrumental variable approach. Our instrument is lightning strikes per square kilometre. The theory underlying our choice of instrument is that Internet penetration is highly sensitive to power disruptions (Agrawal (2021); Manacorda and Tesei (2020); Andersen and Dalgaard (2013); Andersen et al. (2011)). Power surges and falls lead to equipment failure and damage. Additionally, power outages caused by lightning strikes are likely to reduce the productivity of IT capital (or its marginal benefit). A higher frequency of power disruptions implies a higher maintenance cost for Internet providers and less investment. Hence, there will be less Internet usage in these locations (Andersen et al. (2012)). A natural phenomenon that produces power disturbances is lightning activity. Approximately one-third of all power disruptions in the U.S. are a result of lightning strikes (Andersen et al. (2011)). We, therefore, hypothesize that higher lightning density is a viable candidate instrument for broadband Internet diffusion. The exogenous nature of lightning strike density satisfies the exclusion restrictions condition of the instrumental variable approach.[20]

We use the number of lightning strikes per square kilometre with a one-year lag as an instrument for Internet diffusion. This instrument has the potential to generate useful variations in broadband Internet penetration for each following year. Figures 2-1 and 2-2 compare the rates of broadband penetration between 2010 and 2016, and Figure 2-3 depicts the distribution of lightning density in the U.S. at the county level. These maps show that areas with a higher lightning strike density have experienced a slower broadband diffusion.

As of 2019, all 10 states with the highest lightning strike density were also in the bottom 10 states in terms of Internet penetration. Each panel of Figure 2-8 shows that the Internet diffusion rate negatively correlates with the density of lightning strikes. Although the diffusion rate increases yearly, lightning outbreaks and Internet diffusion maintain their negative relationship over time. Our main analysis is based on two-stage least-squares (2SLS) estimation of α_1 in equation 2-E.1, is the second stage estimates. The first stage equation is specified as follows:

$$Inter_{ct} = \beta_0 + \beta_1 LightD_{c(t-1)} + X'_{ict}\delta + \mu_c + \lambda_k + \gamma_t + \theta_i + \epsilon_{ict} \qquad \text{(2-E.2)}$$

Where $LightD_{c(t-1)}$ is the lightning density per square kilometer in county c in year t-1. Using equation 2-E.1 and 2-E.2 jointly, we find the causal impact of broadband penetration on a financial decision, leveraging exogenous variation in lightning density as an instrument for broadband penetration. The validity of our 2SLS estimates relies on the assumption that conditional on the controls and the fixed effect dummies, lightning strike density affects the financial activities of adults in the U.S. only through its impact on broadband penetration. This identification assumption might not hold unconditionally, as lightning strikes may affect individuals' financial activities through other channels as well. We present tests supporting our identification assumption in section 2.7

Columns (1) and (2) of table 2-2 present the first-stage regression of the lag of lightning density on broadband penetration. In column 1, the regression includes all control variables and fixed effects except for individual fixed effects, while column 2 contains all controls and fixed effects. We found the instrument to be a strong predictor of broadband penetration. A one-strike increase in lightning density is associated with a decrease in the broadband Internet penetration index by around 0.0218 index points in the following year. The value of the reported F-statistics, conditional on controls, is larger than 10, and thus suggests that our choice of instrument is valid. In other words, lightning density has the capacity to affect

[20]approximately 83%of all lightning strikes never touched down

broadband diffusion.

2.6 Results

2.6.1 Broadband penetration and stock market participation

Table 2-3 presents the average effect of broadband diffusion on stock market participation between 2010 and 2016. [21] Our results indicate that higher broadband penetration positively affects stock market participation. Specifically, the results in column 3 indicate that a one index point increase in broadband penetration increases the probability of stock market participation by 2.4 percentage points, on average. Column 4 shows the results generated by estimating equation 2-E.1, which includes additional controls for individual fixed effects. After including individual-level fixed effects, the estimated marginal effect of broadband penetration increases to 3.3 percentage points.

In each column, we also control for the vectors of individual and household level control variables discussed in Section 2.5. As we can see from the table, the respondent's gender matters for investment in the stock markets. Women invest less than men in stocks. Also, Hispanic people invest less compared to non-Hispanic individuals. Black African Americans' investment in the stock market is lower than that of white Americans. As could be expected, education also plays a role. Individuals who graduated from college or university have the highest investment probability. Family income and wealth positively and significantly affect stock market participation. Furthermore, married people have a higher probability of investment in the stock market. People who live in a county with a higher mean of stock market participation have a higher likelihood of investment in the stock market. Having a child decreases participation in the stock market. Finally, individual personality traits such as sociability directly affect stock market participation, and depression has a negative im-

[21] One significant drawback of the HRS data is that it focuses exclusively on older individuals and therefore is not a representative sample of the American population. Of particular concern is the fact that older individuals are less likely to invest in the stock market via the use of new technology (Barber and Odean (2002)). This introduces another source of potential bias. However, Fisher et al. (2016) in a study on the generation above 55 years old in the U.S., using data from the Health and Retirement Study collected from 2002–2014, represents that adults living in the technological era have a high level of Internet perception and are more familiar with the technology. Internet usage among older individuals increased over time. In 2002 only 32% of individuals over the age of 55 used the Internet on a regular basis. By 2014, this figure had increased to 61 %. A more recent study by Macdonald and Hülür (2021), using the same data set, confirms that newer cohorts of older adults are more familiar with the Internet and the use of technology in general. In figure 2-5, we can see that Internet adoption and Internet use are increasing among older adults, the same as other age groups. This indicates that our findings may be downwardly biased and should only be interpreted as a lower-bound estimate of the true effect of broadband penetration on stock market participation and other financial behaviours.

pact. The results in table 2-3 corroborate the graphical illustration of the correlation between Internet penetration and stock market participation in figure 2-7(b).

The estimates produced using the instrumental variable (IV) strategy are presented in columns 7 and 8 of table 2-3. The IV results are generally in line with those produced using OLS, though the magnitude of the estimated effect is higher. A one-unit increase in the Internet penetration index is associated with an increase in the probability of stock market participation by 4.7 percentage points. Including individual fixed effects in column 8 reduces the magnitude of this effect to 4.3 percentage points.

2.6.2 Broadband penetration and mortgage ownership

Columns 3 and 4 of Table 2-7 show the results for our baseline OLS estimations which use the probability of having a mortgage as their dependent variable. We find a positive and statistically significant relationship between broadband penetration and the probability of having a mortgage. Column 3 shows the OLS results for the version of this regression that excludes individual fixed effects. The results in Column 3 indicate that a one index point increase in broadband penetration is associated with an average increase in the probability of having a mortgage by 4.1 percentage points. After including individual fixed effects in the regression specification in column 4, this effect drops to 3.1 percentage points. The probability of having an outstanding mortgage increases if the respondent is a woman, has a higher education level, is not depressed, has children, is married, is not retired, has higher income and wealth, and is sociable. Also, black African Americans and Hispanics invest less in housing. Finally, living in a county with a higher mean of housing investment positively affects individual housing investment. The results generated using the instrumental variable approach, shown in columns 7 and 8, also represent a positive relationship between Internet penetration and the probability of owning a mortgage but have a larger magnitude. In the IV specification in column 7, which excludes individual fixed effects, a one-index point increase in broadband penetration increases the probability of having a mortgage by 4.6 percentage points. In the IV specification, which includes individual fixed effects (see column 8), a one index point increase in the broadband penetration index is associated with an increase in the probability of having a mortgage by 4.3 percentage points. Both the OLS and IV estimates confirm that greater access to high-speed Internet increases the probability of having a mortgage.

2.6.3 Broadband penetration and saving behavior

In this section, we test the effect of broadband penetration on the probability of having an active checking or savings account. Table 2-8 shows the results of both the OLS and IV estimations of this effect. The results of the OLS specifications are shown in columns 3 and 4. Column 3 shows the OLS results for the specification which excludes individual fixed effects. The results for Column 3 suggest that a one index point increase in broadband penetration increases the probability of having active savings or checking accounts by 5.3 percentage points. Column 4, shows the results of the OLS specification, which includes individual fixed effects. The Results in column 4 indicate that a one index point increase in Internet penetration increases the probability of having savings or a checking account by 4.8 percentage points. Being Hispanic, black, low educated, depressed, older, or retired has a negative effect on savings. However, being a female, sociable, or richer has a positive effect. Also, living in a county with a higher saving rate increases the possibility of having an active checking or savings account. The IV estimates, presented in columns 5 through 8, confirm Internet penetration's positive and significant impact on savings behavior. Specifically, a one index point increase in broadband Internet penetration increases the probability of having an active checking or savings account by 6.2 percentage points.

2.6.4 Interaction between financial activities

Our results from the aforementioned OLS and IV specifications indicate there is a positive and statistically significant relationship between broadband penetration and the probability of stock market participation, having a mortgage, and having active savings or checking account. In this section, we explore how these financial behaviours affect one another. In panel A of table 2-9, we split our sample into two sub-samples, the first consisting of individuals who have active savings or checking accounts and the other consisting of individuals who do not. From this, we can see that the likelihood of investment in the stock market diminishes significantly for individuals who do not have active savings or checking accounts. Columns 1 to 4 of table 2-9 show that a point increase in the broadband penetration index increases the probability of investment in the stock market by more than 4.3 percentage points among those with active savings or checking accounts, but only increases the probability of stock market participation by 0.02 percentage points among individuals who do not have active savings or checking account. The IV results in columns 5 to 8 show that among individuals with active savings or checking accounts, a one index point increase in Internet penetration is associated with a 6.6 percentage point increase in stock market participation for people with active saving account and 2.5 percentage points for people without active saving or checking account. However, the results in column 8 show that this effect is not significant.

In panel B, we consider the effect of broadband penetration on the probability of having an outstanding mortgage, separately for individuals with and without active savings or checking accounts. For an individual with active savings or checking account, we find that a one index point increase in Internet penetration increases the probability of having a mortgage by 3.2 percentage points (column 2). However, for individuals without an active checking or savings account, we find that the effect of Internet penetration on mortgage ownership is statistically insignificant (columns 3 and 4). The IV results follow a similar pattern. A one index point increase in Internet penetration increases the likelihood of investment in housing by 5.2 percentage points among active savers (column 6). In contrast, the effect is statistically insignificant among individuals without a savings or checking account (column 8).

Panel C of table 2-9 shows how the probability of investment in the stock market is mediated by having an outstanding mortgage. We find that the effect of Internet penetration on investment in the stock market is significantly weaker for individuals with a mortgage. Results from our OLS specification, which includes individual fixed effects, suggest that a one index point increase in Internet penetration increases the probability of investing in the stock market by 3.21 percentage points for individuals without a mortgage (column 4) but only increases the probability of investment in the stock market by 2 percentage points for individuals with a mortgage (column 2). We hypothesize that this is due to mortgage-owning individuals losing part of their purchasing power because of mortgage payments. We also find that this effect is statistically insignificant when we include individual fixed effects (column 2). In the IV specification, a one index point increase in broadband penetration increases the likelihood of investment in the stock market by 5.18 percentage points for non-mortgage owners (column 8) and 2.8 percentage points for mortgage owners (column 6).

Our analysis shows how the interconnection between financial decisions affects in response to broadband Internet penetration. Further, it can explain why participation in the stock market remains low in specific population groups in the U.S., despite the existence of the biggest financial markets in this country.

2.6.5 Heterogeneity analysis

The preceding sections document the causal average effects of broadband Internet on financial decisions. In particular, we have seen that access to broadband Internet has a positive effect on stock market participation. However, it is unlikely that this effect is uniform across the entire population. Instead, the magnitude of this effect may vary between individuals with different demographic and socio-economic characteristics. In this section, we explore this possibility. The results below show that the effect of broadband access on stock market participation is mediated by gender, age, education, income, marital status, sociability status, and retirement status. These findings shed some light on possible mechanisms through which broadband Internet affects stock market participation among U.S. adults.

Effect by gender

This section examines how the effect of broadband penetration on stock market participation is mediated by gender. The literature points to several factors driving a gap between men's and women's participation rates in financial markets. These factors include lower financial literacy among men (Van Rooij et al. (2011); Almenberg and Dreber (2015)), a gender gap in income, and a higher level of risk-aversion among women (Charness and Gneezy (2012); Croson and Gneezy (2009)). In light of these studies, one would expect Internet access to affect the stock market participation rates of men and women differently.

To test this hypothesis, in columns 1 to 4 of table 2-4, we run separate regressions for each gender and find that broadband penetration has a positive and significant effect on both groups. However, the results also indicate that the magnitude of this effect is larger for men than for women. Specifically, we find that a one index point increase in Internet penetration increases stock market participation by 3.6 percentage points for males and 3.1 percentage points for females. The instrument variable regressions show a similar result, though with a higher magnitude. Specifically, one index point increase in Internet penetration increased the stock market participation rate for women by 4.3 percentage points, affecting men's and women's stock market participation rates by 4.89 percentage points. Our results align with the previously discussed literature pertaining to the gender gap in the stock market.

Effect by education level

One would also expect the effect of broadband penetration on stock market participation to vary between individuals with different levels of education. There are several reasons for this. Educated Internet users have lower fixed costs of participation in stock markets, as educated individuals can more easily collect financial information. Additionally, they are more likely to understand the market's risks and rewards and can trade more competently than their less-educated counterparts (Hong et al. (2004); Bayer et al. (2009)). To test how the effect of broadband penetration varies by level of education, we separated the sample into two groups: individuals with a high school level education or below and individuals with a level of education above high school.

In our sample, the stock market participation rate for the more educated group is 27 percent, while the participation rate is around 11 percent for the less educated group. Columns 5, 6, 7, and 8 of table 2-4 shows the OLS and 2SLS results of the effect of broadband penetration on stock market participation for individuals in these two groups.

Results for both the OLS and IV specifications, shown in panels A and B of table 2-4 respectively, indicate that the effect of broadband access on stock market participation is larger for individuals in the more educated group. Results for the OLS specifications indicate that a one index point increase in broadband penetration is associated with an increase in stock market participation by 3.2 percentage points for the more educated group and 2.2 percentage points for the less-educated group. IV estimates indicate that one index point increase in broadband penetration results in a 5.29 percentage point increase in the stock market participation rate for the more educated group and a 4.4 percentage point increase for the less educated group. This suggests that educated individuals can better understand how to use the Internet and invest more in the stock market.

Effect by year of birth cohort

In order to examine the effect of broadband penetration on the stock market participation of different age groups, we split the sample into different age cohorts according to the HRS age segmentation for individuals' years of birth. The OLS and IV results for this analysis are shown in table 2-6. The effect of broadband penetration on stock market participation is highest for the youngest age cohort: the late baby boomers. Members of this cohort were between 51 and 56 years old in 2016. For this age group, a one index point increase in broadband penetration is associated with a 3.55 percentage point increase in stock market participation. The effect of stock market participation for war babies, early baby boomers (EBB), age between 63 to 68 years old and mid-baby boomers (MBB), aged between 57 to 62 years old in 2016, is lower than late BBs. However, the effect remains significant for all age groups except for individuals in CODA and AHEAD age groups that corresponded to people in their 76 to 81 years old and higher in 2016. The IV results are consistent with the OLS estimates, though with a higher magnitude. In the IV specifications, one index point increase in broadband penetration increases stock market participation for the late baby boomers by 5.9 percentage points. Similarly to the OLS specifications, the effect of broadband penetration on stock market participation became statistically insignificant for age groups above 76 years old in 2016.

Effect by marital status

It is understood from the household financial literature that married individuals invest more in the stock market and generally invest in more risky assets compared to single investors (Guiso et al. (2003)). Marriage affects the costs of running a household due to economies of scale (Blow et al. (2009)). Married individuals also have the opportunity to save more. Indeed, there is plenty of evidence that married individuals have higher savings and wealth than unmarried individuals (Zagorsky (2005); Guner and Knowles (2004)). In our dataset, the stock market participation rate is 13 percent for non-married individuals and 26 percent for married individuals.

Results for the OLS and IV estimations, in columns 9 to 12 of table 2-4, show that the effect of Internet access on stock market participation is higher for married individuals than for non-married individuals. In the specification that includes all the controls, a one index point increase in broadband Internet penetration increases stock market participation by 3.2 percentage points for married individuals and 1.3 percentage points for non-married individuals. The IV results follow the same pattern. A one index point increase in Internet penetration increases stock market participation by 4.8 percentage points for married individuals and 2.16 percentage points for non-married individuals. However, the effect is not significant for non-married individuals.

Effect by mood

Behavioural finance and investor sentiment theory establish that investors' mood, such as optimism or pessimism about future market values, affects their stock market participation. The pessimistic feeling is followed by fewer financial activities and lower investment in stock markets, while optimism has the opposite effect (Bollen and Mao (2011); Hong et al. (2004)). We examine how the effect of Internet access on stock market participation differs for depressed or non-depressed individuals. We distinguish individuals who reported having a depressing feeling during the two weeks before the survey and those who did not. Our sample shows a lower rate of stock market participation for individuals who were depressed relative to those who were not depressed.

Columns 13 through 16 of table 2-4 compare the effect of broadband penetration on stock market participation for the depressed and non-depressed groups. In the OLS specifications with all the controls, a one index point increase in Internet penetration was found to increase the stock market participation rate of non-depressed respondents by 3.7 percentage points. By contrast, the results produced a non-significant relationship between Internet penetration and stock market participation for members of the depressed group. Similarly, the IV results indicate that a one-unit increase in broadband penetration is associated with a 5.6 percentage point increase in the stock market participation rate for non-depressed individuals, but the effect is not statistically significant.

Effect by sociability status

Trust and sociability are fundamental concepts in the theory of social capital and household portfolio choice. Individuals' level of social interactions is related to the community characteristics (Aghion and Durlauf (2005); Guiso et al. (2006)), which is representative of local culture. Having a larger number of social connections reduces fixed participation costs through cheaper information and can positively affect stock ownership (Georgarakos and Pasini (2011); Guiso et al. (2004, 2008); Hong et al. (2004)).

We examine how the effect of broadband Internet on stock market participation is mediated by sociability. Columns 1 to 4 of table 2-5 compare the effect of broadband penetration on stock market participation for more and less sociable individuals. We find that the effect of Internet penetration on stock market participation is larger for more sociable individuals. According to the OLS estimates that account for all the controls, a one index point increase in broadband penetration increases stock market participation by 3.8 percentage points for more sociable individuals and decreases by 1.2 percentage points for non-sociable individuals. The IV results are qualitatively similar to those obtained using OLS. In the IV specification, a one index point increase in broadband penetration results in a 5.2 percentage point increase in stock market participation for more sociable individuals and a 3.8 percentage point increase for non-sociable individuals. Our results confirm that having access to the Internet has a complementary role over sociability for stock market participation while Liang and Guo (2015), and Fernández-López et al. (2018), who found that Internet use can serve as a substitute for social capital in determining the probability of stock market participation.

Effect by retirement status

Studies on portfolio choice have not reached a consensus in terms of the direction in which retirement affects the rate of stock market participation. Early retirement can increase saving and portfolio choices after retirement (Farhi and Panageas (2007)). At the same time, loss in working income after retirement might negatively affect investment in the stock market (Cocco et al. (2005)). Additionally, people often re-balance their portfolios away from stocks as they get closer to retirement and exit the stock market after retirement (Fagereng et al. (2017); Addoum (2017)). In our sample, retired individuals were less likely to participate in the stock market than their non-retired counterparts.

We examine how the effect of broadband Internet on stock market participation is mediated by retirement status. Our findings for this investigation are presented in columns 5 to 8 of table 2-5. Our OLS estimates with individual fixed effect imply that a one index point increase in broadband penetration increases stock market participation for non-retired individuals by 3.3 percentage points and by 2.5 percentage points for retired individuals. Instrumental variable results are consistent with the OLS results, though with a larger magnitude. In the IV regressions, a one index point increase in broadband diffusion increases the stock market participation rate by 5.6 percentage points for non-retired individuals and 3.39 percentage points for retired. These findings are consistent with the fact that retired individuals participate less in the stock market in our sample and, therefore less likely to change their investment behavior due to Internet expansion.

Effect by income level

Stock market participation differs between individuals with different income levels. Having a higher income increases the probability of saving and thus investment in more risky assets (Hong et al. (2004)). In order to study how the effect of broadband Internet on stock market participation varies by income level, we split our sample into three income categories: a low-income group, consisting of all individuals with an annual household income less than 25,000 USD; a medium-income group, consisting of all individuals with an annual household income between 25,000 and 100,000 USD; and a high-income group, consisting of all individuals with an annual household income higher than 100,000 USD.

Columns 9 to 14 of table 2-5 report the effect of broadband expansion on stock market participation for the different income groups. As expected, the effect of broadband penetration on stock market participation is larger for high-income individuals. OLS estimates with the individual fixed effects show that a one-unit increase in broadband Internet penetration is associated with an increase in the probability of stock market participation of 3.91 percentage points for high-income individuals, and 3.65 percentage points for medium-income individuals. Internet penetration has no statistically significant effect on stock market participation rates for low-income individuals. Results produced using our instrumental variable approach are similar to the OLS results, though with a larger magnitude. Specifically, a one-unit increase in Internet penetration increases the stock market participation rate by 5.9 percentage points for the high-income group, 4.49 percentage points for the medium-income group, and 3.2 percentage points for the low-income group.

2.6.6 Some possible mechanisms

In the previous sections, we found that broadband Internet penetration directly affects individuals' financial decisions in the U.S. This effect likely occurs because increasing the availability of high-speed Internet leads to an increase in the rate of Internet use, financial literacy, and the use of online financial services. In the sections below, we test these possible mechanisms.

Effect of broadband penetration on Internet use

In this section, we show that increasing broadband Internet penetration increases average Internet use within the study period. In 2010, when the average Internet penetration score was 2.5 out of five, 49 percent of adults in our sample were using the Internet. In 2016, the broadband Internet penetration index reached 3.6 out of five, and 61 percent of adults in the sample were using the Internet. Panel (a) of figure 2-7(a) shows a positive correlation between broadband penetration and Internet use.

Table 2-10 summarizes the OLS and 2SLS estimates of broadband penetration on Internet use. The OLS results in column 4, which include all controls and fixed effects, indicate that a one index point increase in broadband diffusion increases Internet use by around 1.8 percentage points.

The IV estimates in columns 5 to 8 reveal that broadband penetration positively and significantly affects Internet use. Specifically, one index point increase in Internet penetration increases Internet use by 3.9 percentage points. Having established that increasing broadband penetration leads to increased Internet use, we next test the causal impact of broadband penetration on various mechanisms through which Internet use may affect financial behaviour.

Financial literacy

The Internet has broadly replaced traditional media as the primary source of financial information in the last decades (Ratchford et al. (2001); Fisher et al. (2015)). As the use of the Internet became more widespread, consumers' financial behaviours rapidly changed (Carlsson et al. (2017)). The Internet has been described as a one-stop information hub that provides seemingly unlimited up-to-date financial information, making it a valuable information source for financial matters.

In order to investigate the causal impact of broadband Internet penetration on the financial knowledge of respondents, we define the comprehensive financial knowledge index, which ranges in value from 1 to 5. This index is based on the respondents' answers to 5 financial literacy questions in the HRS. The construction of this index is explained in detail in section 2.4.

Table 2-11 shows the effect of broadband Internet penetration on the probability of giving a correct answer to each financial literacy question. Results show that people have significantly higher levels of financial literacy in locations with higher Internet penetration. The first five columns of table 2-11 show the results of the OLS regressions of Internet penetration on the probability of providing a correct answer to each of the five questions. A higher level of broadband Internet penetration was found to increase the probability of answering each of these questions correctly. The IV estimates, reported in columns 6 to 10, are consistent with the OLS results.

Table 2-13 shows the effect of broadband diffusion on the comprehensive financial knowledge index. A one index point increase in broadband penetration was found to increase financial literacy by 0.29 index points in the OLS regression and 1.618 index points in the 2SLS regression. The results confirm that higher broadband penetration is positively associated with financial literacy among adults in the United States.

The use of online services

The Internet is the primary building block of digital financial services, such as online banking. In turn, online banking has made many financial services, available to customers at a low-cost (Crocker and Williams (2022)). Approximately 64.6 percent of U.S. bank customers reported using online banking in 2021 [22] and around 40 percent of respondents in wave 13 of the HRS survey, which corresponded to the year 2016, reported using online banking regularly. The right-hand side panel of table 2-13 represents the OLS and IV estimates of the effect of broadband penetration on financial service use among the respondents. A one index point increase in Internet penetration increases online banking use by 3.17 percentage points in the OLS estimates and 4.34 percentage points in the IV estimates. These results confirm that, among survey respondents, a higher level of broadband Internet penetration positively affects the use of online banking.

Having a financial advisor is another financial service that broadly becomes available with access to the Internet. We used the 2016 round of HRS surveys (wave 13) to examine the effect of broadband penetration on the probability of using financial advisors. [23] Among the respondents who provided an answer, 77 percent reported having financial advisors. As the

[22] Statista, 2022.

[23] In wave 13 of the HRS survey, respondents were asked if they could trust a financial advisor in their financial decisions. However, few respondents answered this question.

older generation has a higher level of wealth, they mostly prefer to have a financial advisor (Agarwal et al. (2009)).

Table 2-13 summarizes the effect of broadband penetration on the use of financial advisors. Increasing broadband penetration by one index point increases the probability of having a financial advisor by 1.2 percentage points in the OLS estimates and 2.4 percentage points in the IV estimates.

2.7 Placebo test: Evidence in support of identification assumptions

In this section, we present evidence in favour of our identification assumption. If our identification strategy is valid, one would expect no correlation between broadband Internet penetration and our outcome variables when broadband infrastructure is unavailable. Using this logic, we test the validity of our identification assumption using a placebo test. We performed this test for the regressions using stock market participation as a dependent variable. As the expansion of broadband Internet was primarily realized after the enactment of the telecommunication act of 1995[24], we expect no association between future of broadband adoption [25] and stock market participation rates prior to 1995. To implement our placebo test, we studied if broadband availability in 2016 affected the stock market participation rate in 1992. In 1992, broadband Internet did not exist, and dial-up home Internet connections were not yet widely adopted. Column 1 of table 2-14 presents the results of this test, generated using our OLS technique. Reassuringly, we find no statistically significant relationship between the 2016 level of broadband penetration and the rate of stock market participation in 1992. Column 2 of table 2-14 presents the IV results for thttps://www.overleaf.com/projecthis test, which also find no significant relationship between the two variables.

We also perform another placebo test to verify the validity of our instrumental variable. We regress broadband penetration on the respondent's contribution to the 401(k) retirement plan.[26] Since contributions to a 401k plan is an investment for retirement, which is part of

[24]The primary goal of the 1995 Telecommunication Act was to deregulate the broadcast market and increase competition in the telecommunication industry. This facilitated the entry of communication businesses into the market, leading to an increased rate of expansion for broadband infrastructure.

[25]Years after 1995.

[26]The 401(k) is a defined contribution plan which is a cash or deferred arrangement. Employees can elect to defer receiving a portion of their salary, which is contributed on their behalf, before taxes, to the 401(k) plan. The employer may match these contributions. There is a dollar limit on the amount an employee may elect to contribute each year. Employees who participate in the 401(k) plan assume responsibility for their retirement income by contributing a part of their salary and, in many instances, directing their investments.

the employee contract, and employees get information about it when they sign the contract, access to the Internet does not directly affect the employee's information and decision for the contribution. Individuals with access to the Internet are more likely to accept the risk of higher gain for the investment in the stock markets and contribute less to a 401k plan. Columns 3 to 6 of table 2-14 show that there is no statistically significant correlation between contribution to a retirement plan and broadband penetration. The IV results also do not show a significant relationship. These results again confirm the validity of our instrument.

We perform an additional validity check for our instrumental variable. The instrumental variable approach in our study relies on the assumption that lightning strikes affect stock market participation only through their impact on broadband Internet penetration. Specifically, we test for correlations between lightning strike density and other potential determinants of stock market participation. The existence of statistically significant correlations between lightning strike density and other variables undermines the validity of our instrumental variable approach. If statistically significant relationships between lightning strike density and these variables exist, the observed relationship between lightning strike density and our dependent variables may not be causal. This would lead to the possibility of a spurious correlation between broadband diffusion and the dependent variables in our IV specifications. Results in columns 1 and 2 of table 2-15 provide estimates of the reduced-form equation, where the dependent variable is now a measure of local economic development. Specifically, we examine two indicators of local economic development: the unemployment rate and GDP growth. Both of these dependent variables are measured at the county level. We find that lightning density is not affected the unemployment rate, while GDP growth is higher in counties with higher lightning strike densities. These results imply that the instrument does not reflect the indirect effect of lightning on other economic variables affecting our dependent variables. This test supports the idea that our instrument satisfies the exclusion restriction assumption necessary for its validity.

We also test whether our results are valid if we use another measurement of broadband penetration that considers the Internet subscribers of broadband with a speed of 200 kbps in at least one direction. This measurement of broadband Internet covers the wide range of services available for Internet users. Columns 5 to 8 of table 2-A.2 represent the positive and higher magnitude of the effect of broadband penetration on stock market participation.

Further, we conduct a robustness analysis that accounts for spatial correlation in the standard errors. Conley (1999) emphasizes that spatially correlated standard errors require corrections and develop a method for implementing such corrections. Conley's standard errors recognize potential dependence based on spatial proximity. Hsiang (2010) and Fetzer (2015) provide commands to implement Conley's proposed corrections. Colella et al. (2019) provides a command which accounts for spatial correlation in instrumental variables regressions.We correct for spatial correlation in both the OLS and IV estimations in Panel B of table 2-A.3.[27] Comparing the estimates in Panel A (which do not control for spatial correlation) and Panel B shows that our conclusions are robust in accounting for spatial correlation.

Finally, we estimate the "reduced form" effects of lightning density on stock market participation using the following reduced form specification:

$$stock_{ct} = \lambda_0 + \lambda_1 LightD_{c(t-1)} + X'_{ict}\delta + \mu_c + \lambda_k + \gamma_t + \theta_i + \epsilon_{ict} \tag{2-G.3}$$

Where λ_1 captures the effect of a lightning strike on stock market participation. Notably, the reduced form model makes no assumptions about the nature of broadband diffusion. Thus the parameter λ_1 has a causal interpretation even if the exclusion restriction does not hold. We report estimates of the effect of lightning density on stock market participation between the years 2010 to 2016 in table 2-16. Results suggest that a marginal increase in lightning density lowers stock market participation by 0.07 percentage points. The results confirm that lightning density is a relevant instrument for broadband diffusion.

2.8 Conclusion

The financial behaviour of individuals and households in an economy plays a central role in economic development. The Internet has made many financial services and information sources available to small investors at lower prices and higher speeds. This paper explores the impact of broadband Internet penetration on financial decisions, including stock market participation, investment in housing, and saving behaviour. The analysis focuses on the 2010 to 2016 period. This period covers the 6 years following the implementation of the FCC's National Broadband Plan in March 2010. This plan aimed to provide Americans with faster and cheaper Internet at the national level. To conduct our analysis, we combine geospatial information on broadband diffusion over the 2010-2016 period with individual-level data on financial decisions obtained from the Health and Retirement Study. Exploiting the gradual expansion of broadband Internet access in the United States, we use OLS regressions with time, county, cohort and individual-level fixed effects to address endogeneity issues related to the nature of Internet diffusion. Additionally, we address remaining endogeneity concerns by using an instrumental variable approach. This approach uses lightning density per square

[27] Estimates in Panel B control for spatial correlation within a distance of 100 kilometres.

kilometre as an instrument. Both identification strategies find that broadband diffusion significantly increases respondents' probability of participating in the stock market, having a mortgage, and having a checking or savings account.

We also investigate how broadband Internet affects the interaction between these financial decisions. We find that having a mortgage weakens the effect of broadband diffusion on stock market participation. We hypothesize that this may be due to mortgage ownership reducing homeowners disposable income, leading to a higher number of mortgage holders opting out of the stock market. We also find that access to more affordable Internet increases individuals' saving rate, and that a person with a higher saving rate is more likely to invest in the housing and stock markets.

Investigating heterogeneities in the effect of broadband Internet on stock market participation represents a more prominent effect for men and individuals who are more educated, younger, more affluent, less depressed, and more sociable. The analysis of the underlying mechanisms shows that broadband diffusion increases Internet take-up rates, financial literacy, and the use of online financial services.

Our analysis suggests that expanding broadband infrastructure can make financial markets and financial services more accessible to small investors, leading to higher levels of financial development and economic growth.

Figure 2-1: Broadband penetration 2010

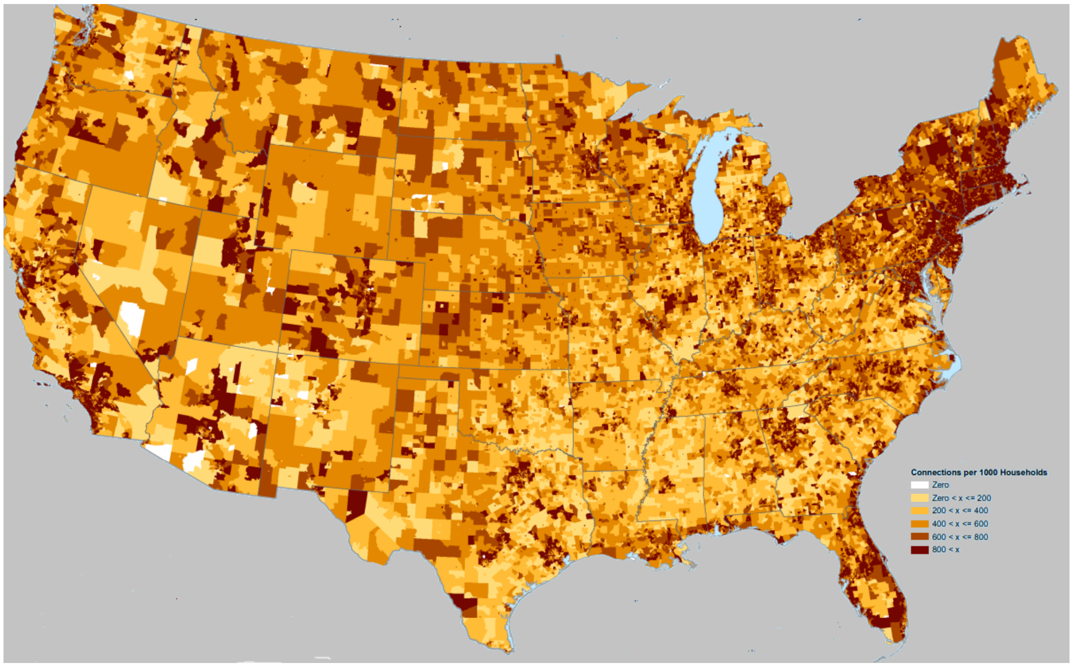
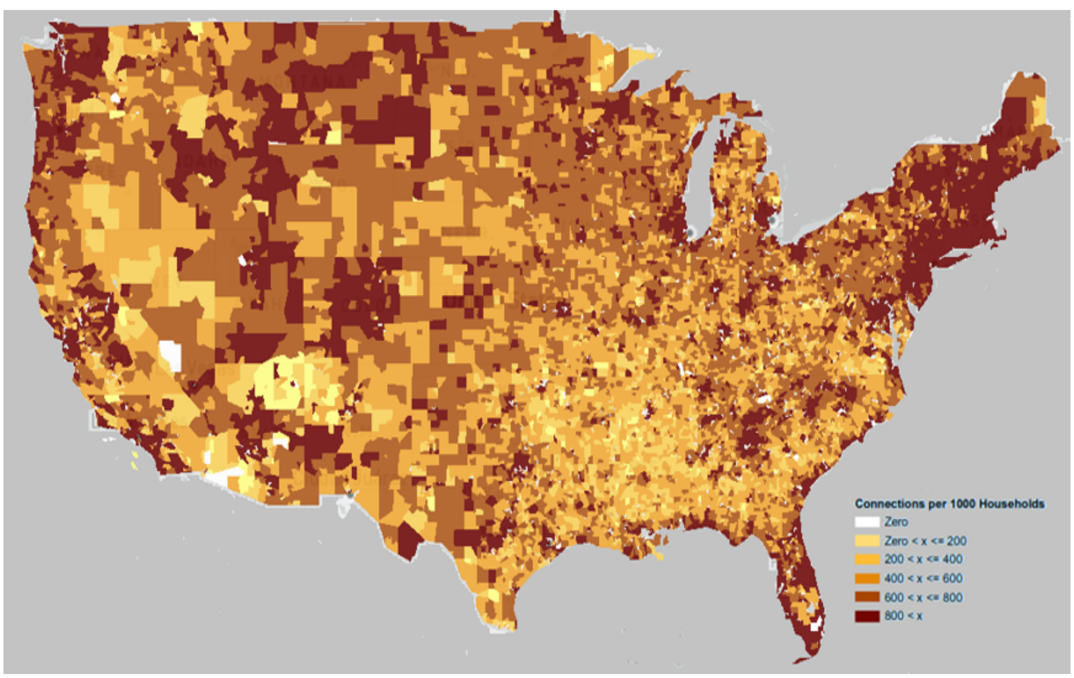


Figure 2-2: Broadband penetration 2016



Note: Each map shows the number of residential fixed Internet access services for Connections over 200 kbps in at least one direction per 1,000 households. Source: Federal Communications Commission.

Figure 2-3: US. average lightning density per square mile between 2008 and 2017

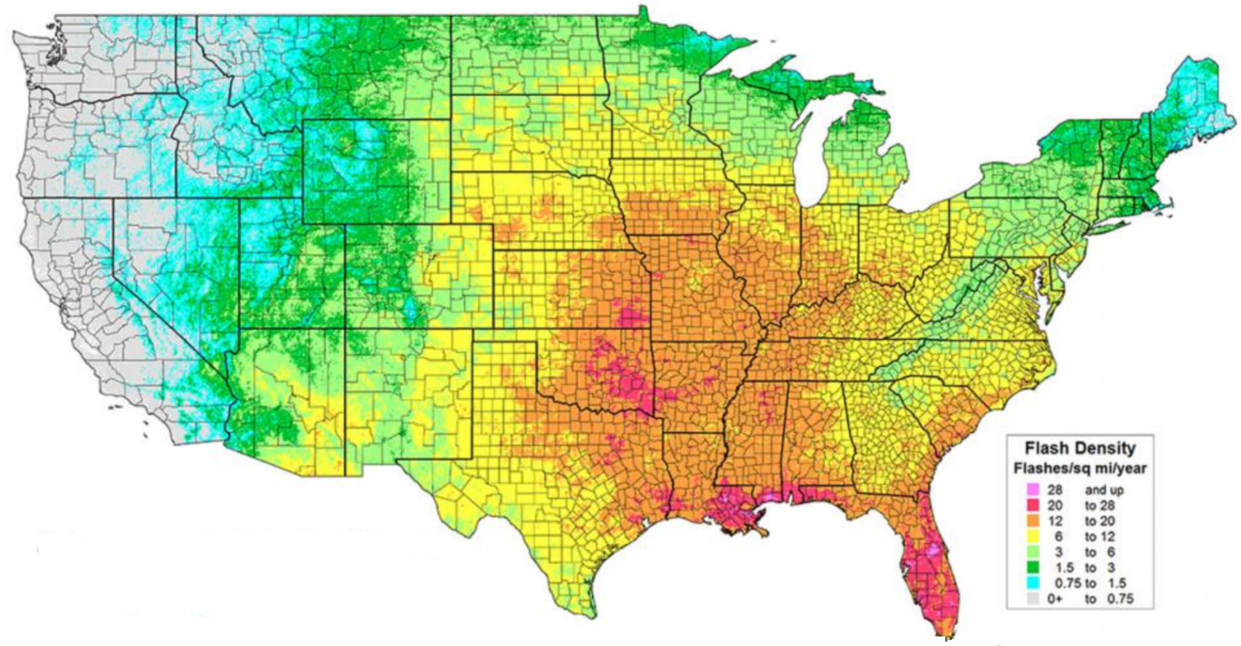
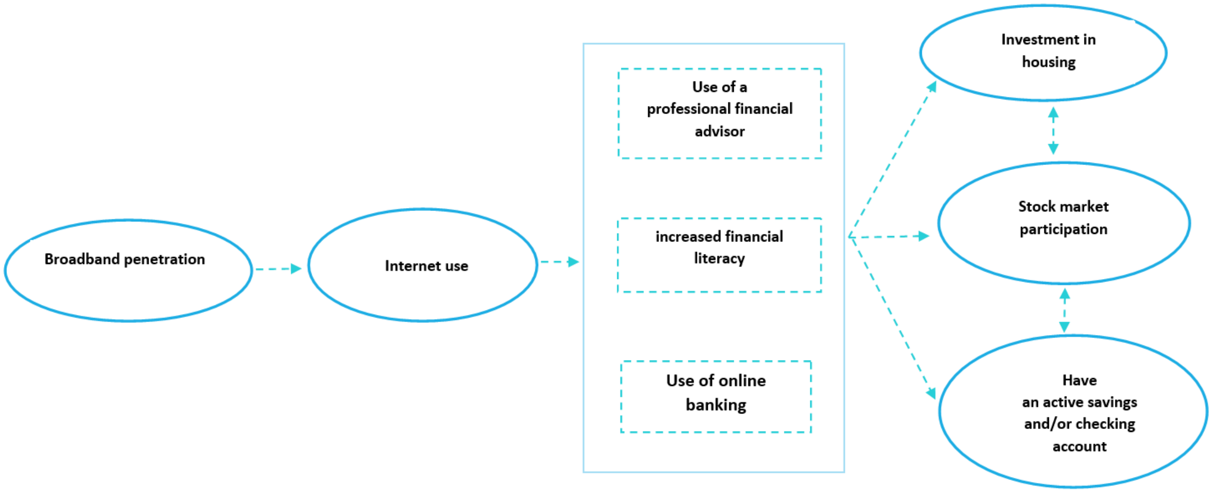
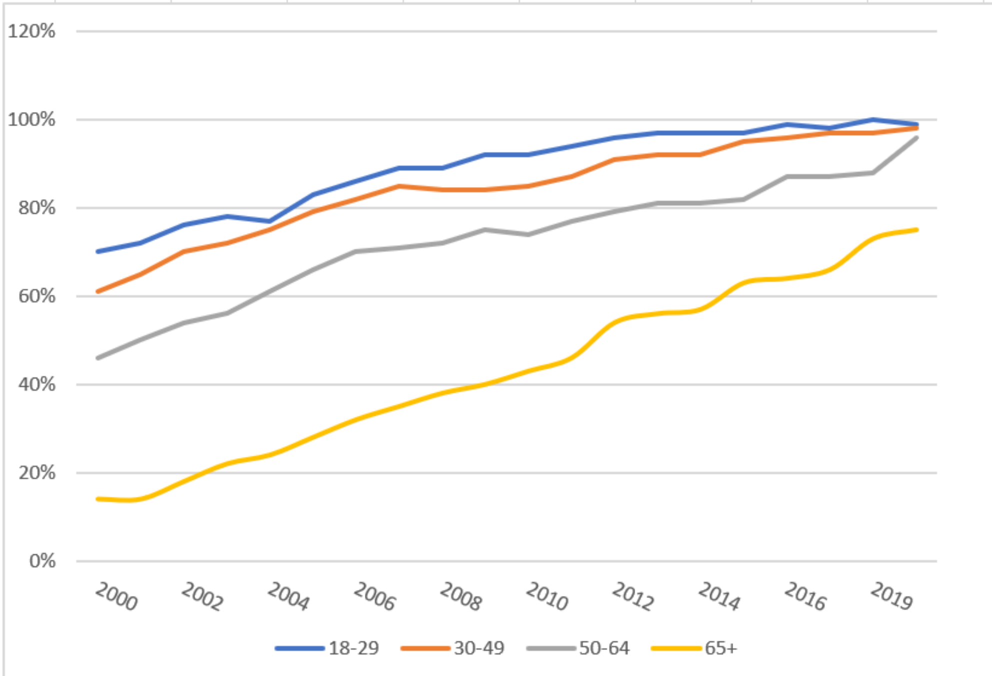


Figure 2-4: Suggestive conceptual framework

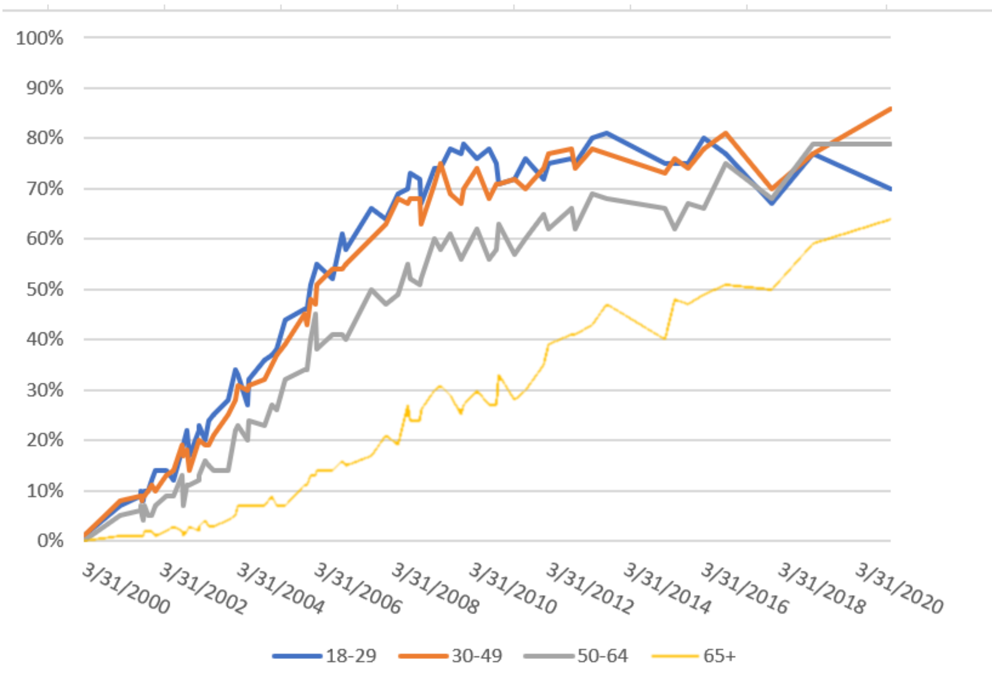


NOTE: The diagram represents a mechanism through that Internet penetration can affect individual financial activities and decisions.

Figure 2-5: Broadband Internet adoption and Internet use by age group in the U.S.



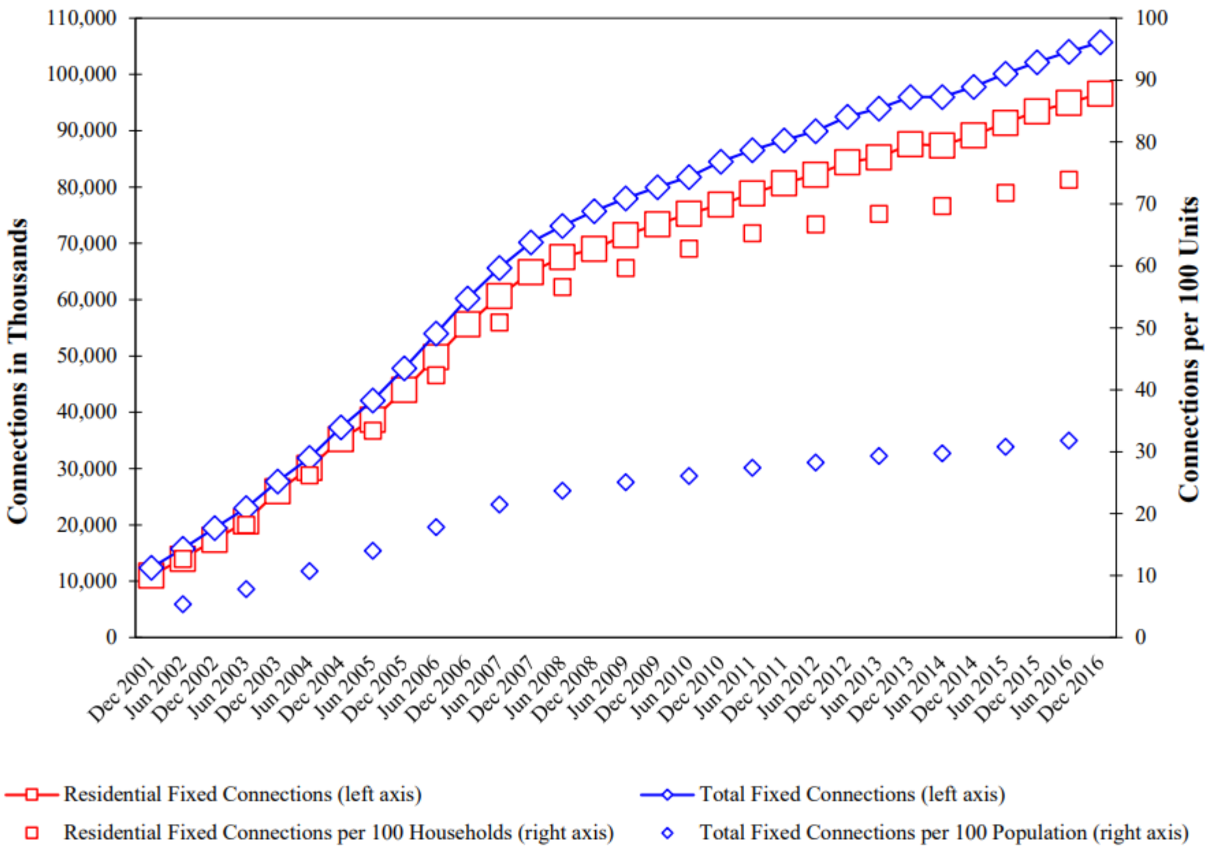
(a) Internet use by different age groups



(b) Have broadband connection at home by age group

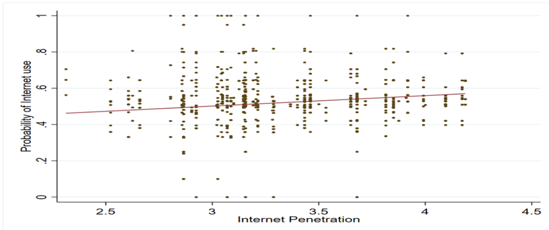
Source: Pew research center, Surveys of U.S. adults conducted 2000-2021. Data for each year is based on a pooled analysis of all surveys conducted during that year.

Figure 2-6: Fixed-location connections 2001-2016

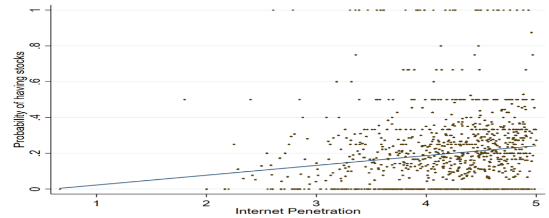


Sources: Internet Access Services: Status as of December 31, 2016.

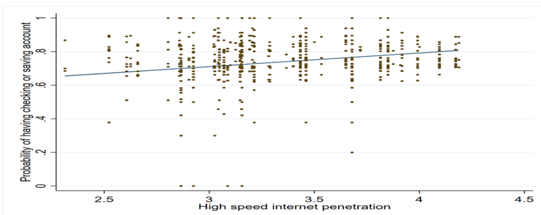
Figure 2-7: Broad band Internet penetration and financial activities



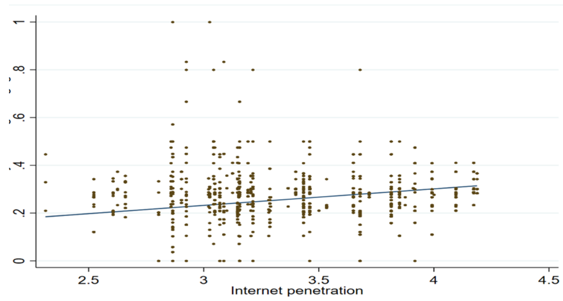
(a) Broadband penetration and probability of Internet use



(b) Broadband penetration and probability of having stocks and mutual funds



(c) Broadband penetration and probability of having an active savings account



(d) Broadband penetration and probability of having mortgage

Figure 2-8: Broadband Internet coverage and lightning

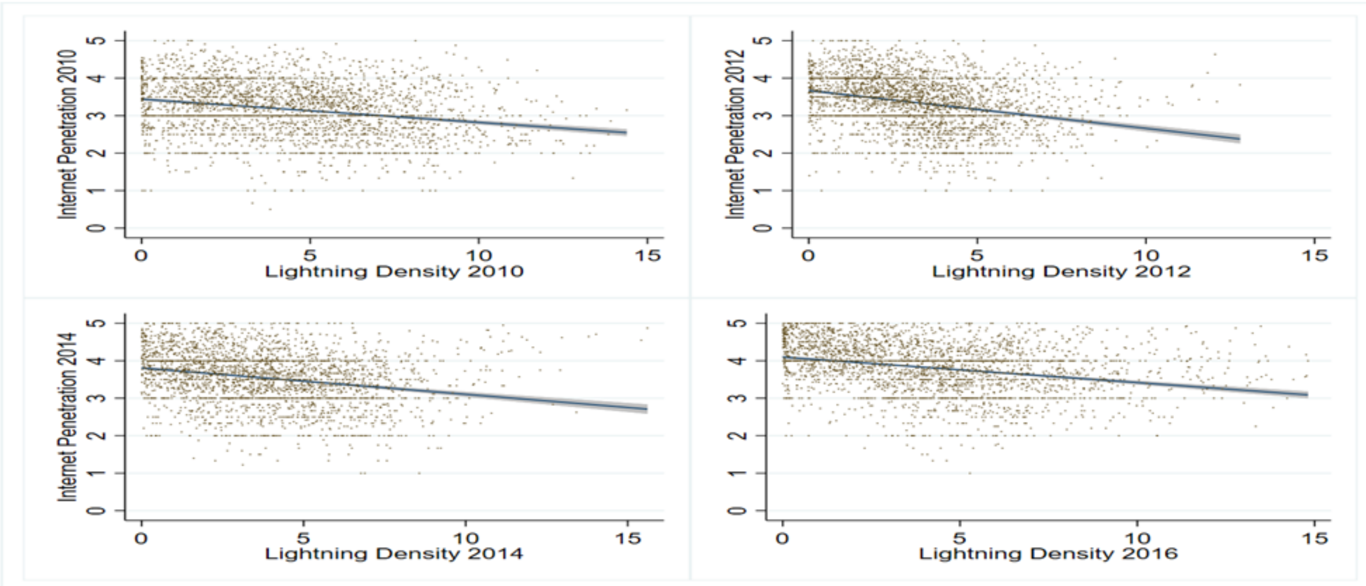


Table 2-1: Summary statistics

	NOB	mean	SD	min	max
Male	8,358	0.415	0.492	0	1
Female	8,358	0.584	0.492	0	1
Have some high school education	8,358	0.197	0.432	0	1
Graduate at high school level	8,358	0.256	0.436	0	1
Have some collage education	8,358	0.290	0.554	0	1
Have education higher than college	8,358	0.252	0.434	0	1
Retired	8,358	0.500	0.488	0	1
Have job	7850	0.347	0.476	0	1
White	8,358	0.658	0.474	0	1
Black African American	8,358	0.241	0.427	0	1
Other race	8,358	0.101	0.301	0	1
Married	8,358	0.461	0.498	0	1
Age	8,358	65.340	11.861	42	97
Age2	8,358	8,358	4,410.008	1,764	8,401
Sociable	8,358	0.726	0.445	0	1
Depressed	8,358	0.139	0.358	0	1
Have child	8,358	0.485	0.499	0	1
Live in same place of residence	8,358	0.450	0.497	0	1
Unemployment at the county level	8,358	7.54	2.74	2.2	29.4
GDP per capita growth	8,358	0.041	0.530	-0.93	14.57
Total income in a year	8,358	68,982.56	110,459	0	2,836,240
Total wealth (thousand dollar)	8,358	363.99	950.6	-1109.5	30,850
Have stocks or mutual funds	7,829	0.193	0.394	0	1
Have stocks in 1992	1,205	0.276	0.441	0	1
Have 401(k) plan	4,501	0.419	0.493	0	1
Have mortgage	8,357	0.301	0.458	0	1
Have checking or savings account	8,358	0.745	0.435	0	1
Use online banking	1,146	0.591	0.384	0	1
Trust financial advisor	723	0.761	0.426	0	1
Correct answer to financial questions :					
Interest rate falls, bond prices will [rise/fall]	455	0.641	0.479	0	1
Pick individual company stocks that have better than average returns	456	0.698	0.459	0	1
With higher inflation rate than interest rate, after 1 year, able to buy more than today	457	0.851	0.356	0	1
Have more money in account after 5years with 2 % interest rate	458	0.885	0.319	0	1
Buying a single company stock usually provides a safer return than a stock mutual fund	384	0.747	0.435	0	1
Comprehensive financial knowledge	384	3.886	1.165	0	5
Internet penetration	8,358	3.385	1.215	0	5
Internet use	7,365	0.615	0.486	0	1
Number of lightning strike per 1km$\hat{2}$	7,119	3.668	3.118	0.000089	15.6
Number of lightning strike per 1m$\hat{2}$	7,119	9.449	10.201	0	28.467

Table 2-2: First stage

	(1)	(2)
Flash strikes density(t-1)	-0.0267***	-0.0218***
	(0.00599)	(0.00529)
Female	0.0527	
	(0.0447)	
Hispanic	-0.0523	
	(0.0871)	
Black African American	-0.179***	
	(0.0593)	
Other race	0.0314	
	(0.0897)	
Age	0.0657**	0.297***
	(0.00224)	(0.0312)
Age^2	-0.000504**	-0.000979***
	(0.000173)	(0.000241)
Have high school	0.189	
	(0.102)	
High school grad	0.309***	
	(0.0721)	
Have some college	0.385***	
	(0.0695)	
Have college or higher	0.616***	
	(0.0808)	
Mean stock participation	-0.111	-0.0457
	(0.0442)	(0.0325)
Same residence	-0.0155	
	(0.0442)	
Depressed	-0.112**	-0.0963**
	(0.0412)	(0.0451)
Have children	0.0141	0.0136
	(0.0339)	(0.0417)
Married	-0.0731*	-0.0716*
	(0.0387)	(0.0359)
Retired	-0.0701**	-0.0700**
	(0.0265)	(0.0267)
Sociable	0.0234	0.0580
	(0.0911)	(0.123)
Household income	0.000304***	0.0000416**
	(0.0000104)	(0.0000152)
Household wealth	0.000295**	0.0000932
	(0.000112)	(0.000185)
Survey-year FE	✓	✓
County FE	✓	✓
Individual FE		✓
Cohort FE	✓	✓
N	7,125	7,125
R2	0.426	0.428
F-statistics, excluded instruments	195.54	187.98

Notes: The dependent variable is the broadband Internet penetration index, changing between zero to five, depending on the broadband Internet lines per 1000 households in each county. The independent variable is the lag of the mean percentage of lightning density per square kilometre at the county level between 2009 and 2015. Column 2 includes county and survey-year fixed effect and all household and individual level control variables that we use in OLS specification to estimate the effect of broadband Internet on financial decisions. In column 2, we add the individual fixed effect. The f-statistic for the validity of the IV is reported across the table. Robust standard errors in parentheses clustered at the county level. * p < 0.1, ** p < 0.05, *** p < 0.001.

Table 2-3: Broadband Internet and probability of having stock or mutual funds

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Broadband penetration	0.0237***	0.0240***	0.0248***	0.0301***	0.0402**	0.0452**	0.0473**	0.0433**
	(0.00442)	(0.00524)	(0.00554)	(0.00761)	(0.0177)	(0.0249)	(0.0252)	(0.0210)
Female	-0.000881	-0.00132	-0.00290		-0.00542	-0.00835	-0.00628	
	(0.0132)	(0.0133)	(0.0136)		(0.0132)	(0.0138)	(0.0163)	
Hispanic	-0.0553**	-0.0554**	-0.0796***		-0.0478**	-0.0433**	-0.0505**	
	(0.0180)	(0.0181)	(0.0204)		(0.0184)	(0.0192)	(0.0175)	
Black African American	-0.0904***	-0.0895***	-0.0987***		-0.0772***	-0.0645***	-0.0624**	
	(0.0137)	(0.0137)	(0.0150)		(0.0152)	(0.0177)	(0.0268)	
Other race	-0.0468**	-0.0468**	-0.0463*		-0.0460**	-0.0491**	-0.0407*	
	(0.0176)	(0.0176)	(0.0184)		(0.0219)	(0.0234)	(0.0211)	
Age	-0.00318	-0.00353	-0.00275	-0.00595	-0.00423	-0.00446	-0.00450	-0.00467
	(0.00615)	(0.00619)	(0.00610)	(0.0107)	(0.00680)	(0.00695)	(0.00610)	(0.00925)
Age^2	0.0000285	0.0000309	0.0000245	0.0000286	0.0000321	0.0000383	0.0000350	0.0000392
	(0.0000444)	(0.0000460)	(0.0000416)	(0.00004665)	(0.0000499)	(0.0000510)	(0.0000499)	(0.0000811)
Have high school	0.000805	0.00182	0.00695		0.00137	0.00489	0.00721	
	(0.0199)	(0.0200)	(0.0214)		(0.0274)	(0.0263)	(0.0322)	
High school grad	0.0333*	0.0341*	0.0472**		0.0337**	0.0285**	0.0290*	
	(0.0140)	(0.0135)	(0.0149)		(0.0129)	(0.0125)	(0.0167)	
Have some college	0.0561***	0.0545***	0.0602***		0.0377**	0.0329**	0.0375*	
	(0.0139)	(0.0134)	(0.0145)		(0.0155)	(0.0170)	(0.0239)	
Have college or higher	0.144***	0.140***	0.149***		0.112***	0.111***	0.0921**	
	(0.0189)	(0.0190)	(0.0195)		(0.0286)	(0.0318)	(0.0323)	
Mean stock participation	0.213***	0.201***	0.189***	0.165***	0.185***	0.193***	0.170***	0.144***
	(0.0425)	(0.0460)	(0.0445)	(0.0303)	(0.0507)	(0.0540)	(0.0500)	(0.0442)
Same residence	0.0217*	0.0215*	0.0212		0.0277*	0.0332*	0.0289	
	(0.0110)	(0.0105)	(0.0155)		(0.0133)	(0.0149)	(0.0211)	
Depressed	-0.0254***	-0.0248**	-0.0253**	-0.0152	-0.0269**	-0.0281**	-0.0244**	-0.0153
	(0.00767)	(0.00767)	(0.00799)	(0.0107)	(0.00876)	(0.00888)	(0.00863)	(0.0109)
Have children	-0.00807	-0.00847	-0.0102	-0.0136	-0.00610	-0.00793	-0.0142	-0.0163
	(0.00845)	(0.00845)	(0.00860)	(0.0110)	(0.00950)	(0.00962)	(0.0113)	(0.0130)
Married	0.0447***	0.0440***	0.0449***	0.0377**	0.0631***	0.0589***	0.0556**	0.0489**
	(0.0117)	(0.0117)	(0.0119)	(0.0139)	(0.0136)	(0.0136)	(0.0174)	(0.0190)
Retired	-0.0314**	-0.0328**	-0.0348**	-0.0331*	-0.0349**	-0.0292**	-0.0250*	-0.0251*
	(0.0113)	(0.0115)	(0.0117)	(0.0177)	(0.0147)	(0.0136)	(0.0144)	(0.0129)
Sociable	0.127***	0.127***	0.128***	0.125***	0.129***	0.126***	0.121***	0.116***
	(0.00899)	(0.00906)	(0.00939)	(0.0129)	(0.0116)	(0.0110)	(0.0130)	(0.0150)
Total income	0.000316***	0.000314***	0.000315***	0.000241**	0.000247***	0.000244***	0.000233**	0.000226**
	(0.0000705)	(0.0000707)	(0.0000713)	(0.0000811)	(0.0000724)	(0.0000731)	(0.0000837)	(0.0000889)
Total wealth	0.00436***	0.00435***	0.00425***	0.00189**	0.00368***	0.00371***	0.00275***	0.00169**
	(0.00120)	(0.00119)	(0.00115)	(0.000862)	(0.00103)	(0.00104)	(0.000759)	(0.000765)
Survey-year FE		✓	✓	✓		✓	✓	✓
County FE			✓	✓			✓	✓
Individual FE				✓				✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
N	6,631	6,631	6,631	6,631	6,631	6,631	6,631	6,631
R2	27.88	27.20	29.43	21.11	22,49	20.41	24,51	18.88

Notes: The table reports the OLS and 2SLS estimates of the effect of broadband penetration on the probability of having stocks or mutual funds. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-4: Heterogeneous effects of connecting to the Broadband Internet on the probability of having stock or mutual funds

	Women		**Men**		**High education**		**Low education**		**Married**		**Not- married**		**Not -depressed**		**Depressed**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Panel A (OLS)															
Broadband penetration	0.0258***	0.0318***	0.0262***	0.0361**	0.0377***	0.0325***	0.0192**	0.0226**	0.0284***	0.0327***	0.0125**	0.0133*	0.0259***	0.0370***	-0.0107	0.0149
	(0.00681)	(0.00954)	(0.00656)	(0.0125)	(0.00606)	(0.00847)	(0.00762)	(0.0125)	(0.00624)	(0.00873)	(0.00533)	(0.00771)	(0.00559)	(0.00812)	(0.00872)	(0.0235)
N	3,911	3,911	2,720	2,720	5,334	5,334	1,297	1,297	2,998	2,998	3,633	3,633	5,727	5,727	903	903
	Panel B (IV)															
Broadband penetration	0.0451**	0.0430**	0.0492***	0.0489**	0.0529**	0.0527**	0.0371*	0.0441	0.0583**	0.0487**	0.0286*	0.0216	0.0597**	0.0568**	0.0120	0.0218
	(0.0220)	(0.0205)	(0.0133)	(0.0218)	(0.0307)	(0.0268)	(0.0222)	(0.0335)	(0.0267)	(0.0215)	(0.0165)	(0.0182)	(0.0258)	(0.0235)	(0.0212)	(0.0221)
N	3,911	3,911	2,720	2,720	5,334	5,334	1,297	1,297	2,998	2,998	3,633	3,633	5,727	5,727	903	903
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE		✓		✓		✓		✓		✓		✓		✓		✓

Notes: The table reports heterogeneous effects of broadband penetration on the probability of having stocks or mutual funds by gender, education level, marital status, and mood. Panel A shows OLS estimates, and Panel B shows 2SLS estimates. Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth and the number of children. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-5: Heterogeneous effects of connecting to the Broadband Internet on the probability of having stock or mutual funds

	Sociable		**Not- sociable**		**Retired**		**Not-retired**		**Highe level income**		**Medium level income**		**Low level income**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Panel A (OLS)							
Broadband penetration	0.0293***	0.0389***	-0.0101	-0.0127	0.0244***	0.0257**	0.0269**	0.0331***	0.0265***	0.0391***	0.0229**	0.0365***	0.0121*	0.0152
	(0.00660)	(0.00941)	(0.00657)	(0.00699)	(0.00669)	(0.00976)	(0.00770)	(0.0123)	(0.00627)	(0.0115)	(0.00740)	(0.0120)	(0.00733)	(0.0126)
N	4,833	4,833	1,798	1,798	3,325	3,325	3,306	3,306	1,264	1,264	3,002	3,002	2,338	2,338
							Panel B (IV)							
Broadband penetration	0.0510**	0.0523**	0.0327	0.0366	0.0283*	0.0341*	0.0530**	0.0567**	0.0564**	0.0594***	0.0421**	0.0447*	0.0383*	0.0325
	(0.0227)	(0.0235)	(0.0246)	(0.0290)	(0.0131)	(0.0162)	(0.0227)	(0.0244)	(0.0185)	(0.0151)	(0.0192)	(0.0237)	(0.0188)	(0.0211)
N	4,833	4,833	1,798	1,798	3,325	3,325	3,306	3,306	1,264	1,264	3,002	3,002	2,338	2,338
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE		✓		✓		✓		✓		✓		✓		✓

Notes: The table reports heterogeneous effects of broadband penetration on the probability of having stocks or mutual funds for subgroups by sociability, retirement status, and income level. Panel A reports OLS estimates and Panel B reports 2SLS estimates. Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth, and the number of children. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-6: Heterogeneous effects of broadband penetration on the probability of having stocks and mutual funds by year of birth

	AHEAD	CODA	HRS	War	Early BB	Mid BB	Late BB
Panel (A): OLS with cohort, survey-year and county fixed effects							
Broadband penetration	0.00137	0.0121	0.0146*	0.0233***	0.0285***	0.0351***	0.0365***
	(0.00843)	(0.00771)	(0.00741)	(0.00524)	(0.00651)	(0.00667)	(0.00659)
N	406	531	689	1,111	1,615	1,842	437
Panel (B): OLS with the cohort, survey-year, county, and individual fixed effects							
Broadband penetration	0.00724	0.0141	0.0193	0.0248***	0.0301***	0.0375***	0.0392***
	(0.0124)	(0.00944)	(0.00827)	(0.00631)	(0.00771)	(0.00769)	(0.00765)
N	406	531	689	1,111	1,615	1,842	437
Panel (C): IV with cohort survey-year and county fixed effects							
Broadband penetration	0.0125	0.0184	0.0216	0.0433**	0.0439**	0.0548**	0.0591**
	(0.0238)	(0.0146)	(0.0252)	(0.0264)	(0.0202)	(0.0221)	(0.0205)
N	406	531	689	1,111	1,615	1,842	437
Panel (D): IV with the cohort, Survey-year, county, and individual fixed effects							
Broadband penetration	0.00856	0.0149	0.0243	0.0329*	0.0466**	0.0546**	0.0609**
	(0.0259)	(0.0215)	(0.0223)	(0.0194)	(0.0224)	(0.0219)	(0.0235)
N	406	531	689	1,111	1,615	1,842	437
Individual level controls	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓

Notes: Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-7: Broadband Internet and probability of investing in housing

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Broadband penetration	0.0265***	0.0388***	0.0417***	0.0315***	0.0432**	0.0427**	0.0461*	0.0436***
	(0.00443)	(0.00525)	(0.00547)	(0.00689)	(0.0148)	(0.0233)	(0.0250)	(0.0147)
Female	0.0375**	0.0361**	0.0329**		0.0337**	0.0316**	0.0241**	
	(0.0131)	(0.0131)	(0.0133)		(0.0146)	(0.0154)	(0.0102)	
Hispanic	-0.0676**	-0.0659**	-0.0625**		-0.0614**	-0.0525**	-0.0629*	
	(0.0237)	(0.0236)	(0.0251)		(0.0278)	(0.0258)	(0.0392)	
Black African American	-0.109***	-0.106***	-0.115***		-0.102***	-0.0884***	-0.0873**	
	(0.0167)	(0.0167)	(0.0177)		(0.0192)	(0.0215)	(0.0359)	
Other race	-0.0711**	-0.0722**	-0.0739**		-0.0746**	-0.0819**	-0.0787**	
	(0.0243)	(0.0243)	(0.0250)		(0.0286)	(0.0305)	(0.0396)	
Age	-0.00571	-0.00430	-0.00442	-0.0212*	-0.00660	-0.00690	-0.00722*	-0.0312**
	(0.000731)	(0.000728)	(0.000742)	(0.0127)	(0.000600)	(0.000745)	(0.00361)	(0.00126)
Age^2	0.0000239	0.0000262	0.0000299	0.000105	0.0000725	0.0000770	0.0000812	0.000980**
	(0.0000551)	(0.0000553)	(0.0000565)	(0.0000944)	(0.0000610)	(0.0000650)	(0.0000570)	(0.0000321)
Have high school	0.0559*	0.0543*	0.0649**		0.0443	0.0469*	0.0635**	
	(0.0285)	(0.0263)	(0.0291)		(0.0320)	(0.0247)	(0.0268)	
High school grad	0.0171	0.0147	0.0170		0.000971	0.00620	0.00410	
	(0.0184)	(0.0184)	(0.0193)		(0.0235)	(0.0270)	(0.0279)	
Have some college	0.0865***	0.0827***	0.0867***		0.0559**	0.0441**	0.0340**	
	(0.0188)	(0.0187)	(0.0195)		(0.0256)	(0.0206)	(0.0120)	
Have college or higher	0.137***	0.131***	0.142***		0.0831**	0.0612**	0.0597**	
	(0.0222)	(0.0221)	(0.0229)		(0.0355)	(0.0242)	(0.0227)	
Mean investment in housing	0.218***	0.221***	0.0952**	0.0871**	0.201***	0.144**	0.0841*	0.0806*
	(0.0653)	(0.0658)	(0.0416)	(0.0355)	(0.0418)	(0.0867)	(0.0428)	(0.0397)
Same residence	0.0769***	0.0778***	0.0761**		0.0884***	0.0858***	0.0857***	
	(0.0163)	(0.0163)	(0.0167)		(0.0183)	(0.0215)	(0.0239)	
Depressed	-0.0645***	-0.0643***	-0.0639***	-0.0572***	-0.0893**	-0.0728*	-0.0927**	-0.0785**
	(0.0109)	(0.0110)	(0.0112)	(0.0134)	(0.0339)	(0.0425)	(0.0393)	(0.0366)
Have children	0.00832	0.00715	0.00636	0.00285	0.0101	0.00990	0.00554	-0.00734
	(0.00929)	(0.00926)	(0.00937)	(0.0117)	(0.0106)	(0.0112)	(0.0140)	(0.0181)
Married	0.169***	0.166***	0.165***	0.0909***	0.175***	0.177***	0.171***	0.106**
	(0.0137)	(0.0137)	(0.0139)	(0.0259)	(0.0151)	(0.0154)	(0.0192)	(0.0368)
Retired	-0.0462***	-0.0389**	-0.0398**	-0.0167	-0.0611***	-0.0406**	-0.0403**	-0.0484*
	(0.0129)	(0.0131)	(0.0134)	(0.0186)	(0.0170)	(0.0151)	(0.0185)	(0.0244)
Sociable	0.0411***	0.0388***	0.0385***	0.0222*	0.0516***	0.0492***	0.0564***	0.0358*
	(0.00976)	(0.00977)	(0.00989)	(0.0118)	(0.0124)	(0.0125)	(0.0167)	(0.0181)
Total income	0.000241**	0.000240**	0.000233**	0.000743**	0.000200**	0.000248**	0.000228**	0.0000654*
	(0.0000785)	(0.0000783)	(0.0000782)	(0.0000357)	(8.31e-08)	(9.15e-08)	(0.000107)	(0.0000332)
Total wealth	-0.0000789	-0.0000924	-0.0000976	-0.000702	-0.000662	-0.00191	-0.000171	-0.000118
	(0.000635)	(0.000618)	(0.000629)	(0.000609)	(0.000633)	(0.00102)	(0.00116)	(0.00181)
Survey-year FE		✓	✓	✓		✓	✓	✓
County FE			✓	✓			✓	✓
Individual FE				✓				✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
N	7,119	7,119	7,119	7,119	7,119	7,119	7,119	7,119
R2	28.11	28.58	31.25	21.73	25.44	23.52	20.32	16.13

Notes: The table reports the OLS and 2SLS estimates of the effect of broadband penetration on the probability of having a mortgage to invest in housing. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-8: Broadband penetration and probability of having active checking or saving account

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Broadband penetration	0.0435***	0.0501***	0.0536***	0.0482***	0.0587***	0.0633***	0.0640***	0.0625**
	(0.00567)	(0.00621)	(0.00641)	(0.00908)	(0.0127)	(0.0163)	(0.0150)	(0.0267)
Female	0.00381	0.00404	0.00750		0.00584	0.00492	0.00648	
	(0.00917)	(0.00918)	(0.00907)		(0.0118)	(0.0170)	(0.0179)	
Hispanic	-0.0935***	-0.0939***	-0.102***		-0.0865***	-0.0867**	-0.0781**	
	(0.0208)	(0.0208)	(0.0211)		(0.0257)	(0.0338)	(0.0363)	
Black African American	-0.112***	-0.113***	-0.100***		-0.0860***	-0.0608**	-0.0571*	
	(0.0128)	(0.0128)	(0.0131)		(0.0164)	(0.0235)	(0.0273)	
Other race	-0.0340*	-0.0339*	-0.0337*		-0.0429*	-0.0762**	-0.0485*	
	(0.0198)	(0.0198)	(0.0196)		(0.0255)	(0.0361)	(0.0275)	
Age	-0.0172**	-0.0185**	-0.0175**	-0.00906	-0.0180**	-0.0210**	-0.0222**	-0.0128*
	(0.00632)	(0.00651)	(0.00663)	(0.0115)	(0.00600)	(0.00745)	(0.00861)	(0.00700)
Age^2	0.000105**	0.000103**	0.0000928*	0.0000826	0.000120**	0.000132**	0.000110**	0.000971*
	(0.0000482)	(0.0000491)	(0.0000499)	(0.0000750)	(0.0000501)	(0.0000511)	(0.0000500)	(0.0000493)
Have high school	0.120***	0.119***	0.120***		0.132	0.131	0.137	
	(0.0350)	(0.0351)	(0.0359)		(0.0353)	(0.0372)	(0.0365)	
High school grad	0.165***	0.165***	0.167***		0.167***	0.171***	0.173***	
	(0.0264)	(0.0264)	(0.0269)		(0.0310)	(0.0315)	(0.0335)	
Have some college	0.228***	0.229***	0.237***		0.291***	0.282***	0.315***	
	(0.0252)	(0.0253)	(0.0254)		(0.0267)	(0.0291)	(0.0312)	
Have college or higher	0.234***	0.235***	0.245***		0.265***	0.264***	0.352***	
	(0.0263)	(0.0263)	(0.0266)		(0.0270)	(0.0254)	(0.0336)	
Mean saving and checking account	0.156***	0.164***	0.0982*	0.121**	0.135**	0.1508**	0.0880*	0.116**
	(0.0451)	(0.0428)	(0.0634)	(0.0573)	(0.0561)	(0.0412)	(0.0487)	(0.0422)
Same residence	0.0294**	0.0296**	0.0264*		0.0323**	0.0385**	0.0341*	
	(0.0137)	(0.0137)	(0.0141)		(0.0155)	(0.0143)	0.0175	
Depressed	-0.242***	-0.242***	-0.241***	-0.238***	-0.117***	-0.0822**	-0.0964***	-0.0438*
	(0.0167)	(0.0167)	(0.0167)	(0.0176)	(0.0255)	(0.0431)	(0.0503)	(0.0233)
Have children	0.00841	0.00854	0.00870	0.0109	0.0154	0.0157	0.0124	0.0339
	(0.00735)	(0.00735)	(0.00742)	(0.0106)	(0.00998)	(0.0129)	(0.0141)	(0.0211)
Married	0.00400	0.00340	0.00284	0.0314	0.0101	0.0161	0.0220	0.0195
	(0.00916)	(0.00917)	(0.00908)	(0.0198)	(0.0124)	(0.0162)	(0.0175)	(0.0288)
Retired	-0.00900	-0.0117	-0.0121	-0.0184	-0.0290****	-0.0537***	-0.0312**	-0.0243**
	(0.0101)	(0.0103)	(0.0103)	(0.0150)	(0.0137)	(0.0169)	(0.0181)	(0.0121)
Sociable	0.143***	0.144***	0.146***	0.140***	0.182***	0.193***	0.190***	0.156***
	(0.0118)	(0.0120)	(0.0121)	(0.0155)	(0.0153)	(0.0174)	(0.0192)	(0.0237)
Total income	0.000325***	0.000301***	0.000411***	0.000533***	0.0000511*	0.000161**	0.000178**	0.000122**
	(0.0000294)	(0.00002882)	(0.0000293)	(0.0000341)	(0.0000308)	(0.0000574)	(0.0000843)	(0.000068)
Total wealth	0.000128	0.000110	0.000213	0.000665	0.000422	0.00150*	0.00140	0.00100
	(0.000202)	(0.000204)	(0.000191)	(0.000365)	(0.000408)	(0.000655)	(0.000796)	(0.00144)
Survey-year FE		✓	✓	✓		✓	✓	✓
County FE			✓	✓			✓	✓
Individual FE				✓				✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
N	7,118	7,118	7,118	7,118	7,118	7,118	7,118	7,118
R2	36.32	36.12	32.09	47.81	26.75	28.80	32.66	34.40

Notes: The table reports the OLS and 2SLS estimates of the effect of broadband penetration on the probability of having a saving account. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-9: Broadband penetration and interconnection between financial decisions

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (A): Investing in the stock market and having an active saving account								
Broadband penetration	0.0324***	0.0438***	0.00671	0.00279	0.0536**	0.0669**	0.0225	0.0253
	(0.00885)	(0.00933)	(0.00732)	(0.0116)	(0.0227)	(0.0243)	(0.0224)	(0.0246)
Have active saving account	✓	✓			✓	✓		
N	4,952	4,952	1,678	1,678	4,952	4,952	1,678	1,678
Panel (B): Investing in housing and having an active saving account								
Broadband penetration	0.0419***	0.0326***	0.0221	0.0217	0.0529**	0.0523**	0.0361	0.0281
	(0.00613)	(0.00794)	(0.00841)	(0.0145)	(0.0226)	(0.0221)	(0.0299)	(0.0258)
Have active saving account	✓	✓			✓	✓		
N	5,229	5,229	1,809	1,809	5,229	5,229	1,809	1,809
Panel (c): Investing in the stock market and housing								
Broadband penetration	0.0227*	0.0206	0.0264***	0.0322***	0.0335*	0.0285	0.0552**	0.0518**
	(0.0129)	(0.0156)	(0.00577)	(0.00842)	(0.0191)	(0.0258)	(0.0223)	(0.0235)
Invest in housing	✓	✓			✓	✓		
N	2,047	2,047	4,583	4,583	2,047	2,047	4,583	4,583
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE		✓		✓		✓		✓

Notes: The table reports the OLS and 2SLS estimates of the effect of broadband penetration on the interaction between financial decisions. Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth and the number of children. The first four columns of panel A show estimates for individuals who participate in the stock market and have an active savings account, and columns 5 to 8 show estimates for individuals who participate in the stock market but do not have an active savings account. Panel B is similar to Panel A by replacing stock market participation with investment in housing. Panel C is similar to Panel A by replacing ownership of active savings accounts with housing investments. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-10: Broadband penetration and probability of Internet use

	OLS				IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Broadband penetration	0.0211***	0.0200***	0.0211***	0.0180**	0.0428**	0.0372**	0.0410**	0.0396**
	(0.00460)	(0.00533)	(0.00580)	(0.00719)	(0.0155)	(0.0156)	(0.0185)	(0.0173)
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE		✓	✓	✓		✓	✓	✓
County FE			✓	✓			✓	✓
Individual FE				✓				✓
N	6,265	6,265	6,265	6,265	6,265	6,265	6,265	6,265

Notes: The table reports the OLS and 2SLS estimates of the effect of broadband penetration on the probability of Internet use. In columns one and five, we control for individual and household level controls, and we add survey-year, county and individual fixed effects in columns 3,4,7, and 8 are based on equations 1,2 and 3. We do not report the individual and household level controls in this table but keep them in each column. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-11: Internet penetration and financial literacy

	OLS					IV				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
High speed Internet penetration	0.0612**	0.0634**	0.0423**	0.0303**	0.0498**	0.0576**	0.0820**	0.0416*	0.0325**	0.0381
	(0.0242)	(0.0253)	(0.0179)	(0.0178)	(0.0211)	(0.0253)	(0.0360)	(0.0214)	(0.0118)	(0.0222)
N	383	383	383	320	383	383	383	383	320	383
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports the effect of broadband penetration on individuals' financial literacy. The first five columns present the OLS estimates for each question. In column 1, individuals were asked whether the bond price would rise or fall. In the second column, they were asked whether it is better to pick a company's stock with a higher than average return. In question 3, they were asked if people have 100 in saving accounts and an interest rate is 2 percent, after five years, will they have more than, equal, or less than 102? Question 4 asks if the inflation rate is more than the interest rate after one year, your money can buy equal, less or more than today. Finally, question 5 asks whether a single company's stock provides a safer return than a stock of mutual funds. The last five-column provide the IV estimates of each question. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-12: Broadband penetration and comprehensive financial knowledge

	(OLS)	(IV)
Broadband penetration	0.343***	1.618**
	(0.0668)	(0.718)
N	320	320
Individual level controls	✓	✓
household level controls	✓	✓
County FE	✓	✓
Cohort FE	✓	✓

Standard errors in parentheses,* $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

Table 2-13: Broadband penetration and financial service use

	Having financial advisor		Online banking	
	(OLS)	(IV)	(OLS)	(IV)
Broadband penetration	0.0122***	0.0241**	0.0317**	0.0434*
	(0.00144)	(0.00945)	(0.0139)	(0.0222)
Individual level controls	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓
County FE	✓	✓	✓	✓
N	576	823	992	823
R2	52.55	40.45	52.27	54.44

Notes: This table presents the OLS and 2SLS estimates of the effect of broadband penetration on the probability of using online services such as having a financial advisor or online banking. For having a financial advisor, we define a dummy variable equal to one if people trust a financial advisor in their financial decisions and zero if not. Also, we define a dummy variable equal to one if people have online banking. Individual level controls are Gender, race, age, age^2, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth, and the number of children. To control for cohort and age effects, we control for cohort fixed effects. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-14: Falsification test for the effect of broadband penetration on stock market participation

	Have stocks in1992		Have 401k retirement plan			
	OLS	IV	OLS		IV	
Broadband penetration	-0.009	-0.0154	0.00147	0.00675	-0.0231	-0.0220
	(0.0103)	(0.0203)	(0.00781)	(0.00835)	(0.0426)	(0.0214)
N	977	977	3,853	3,853	3,853	3,853
Individual level controls	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓
Individual FE				✓		✓

Notes: This table presents the OLS and 2SLS estimates of the effect of broadband penetration in 2016 on the probability of having stocks or mutual funds in 1992, years when broadband was not available for the households in the U.S. and contribution in a 401k retirement plan. Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth and the number of children. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-15: Lightning density and local economy

	Unemployment rate	GDP growth
Lightning density	-0.00880	0.00138***
	(0.00190)	(0.000397)
Survey-year FE	✓	✓
County FE	✓	✓
N	32,104	31,720
R2	1.20	13.35

Notes: This table presents the correlation between lightning density and the local economic situation. Column 1 considers the effect on the unemployment rate between the years 2010 to 2016 at the county level. Data for the unemployment rate is from the U.S. bureau of labour statistics (https://www.bls.gov/lau/). Column 2 presents the impact of lightning on GDP growth during the same period at the county level. We obtain data for GDP growth from SSTI (https://ssti.org/blog/useful-stats-capita-county-level-gdp). In each column, we consider year and county fixed effects. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-16: Effect of lightning density on stock market participation (reduced form)

	(1)	(2)
Lightning density	-0.00799***	-0.00928***
	(0.00299)	(0.00354)
Individual level controls	✓	✓
Household level controls	✓	✓
Cohort FE	✓	✓
Survey-year FE	✓	✓
County FE	✓	✓
Individual FE		✓
N	9,802	9,802

Notes: Table presents the results for the reduced effect of broadband penetration on stock market participation. We directly estimate the effect of lighting on stock market participation according to equation 3. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

2-A Appendix

Table 2-A.1: variable description and data use

	Sample size	Waves	Variable description
Male	38,245	10, 11,12,13	(0, 1 dummy variable)
Female	38,245	10, 11,12,13	(0, 1 dummy variable)
Education:			
Have some high school education	38,245	10, 11,12,135	(0, 1 dummy variable)
Graduate at high school level	38,245	10, 11,12,135	(0, 1 dummy variable)
Have some collage education	38,245	10, 11,12,135	(0, 1 dummy variable)
Have education higher that collage	38,245	10, 11,12,135	(0, 1 dummy variable)
Working status:			
Retired	38,245	10, 11,12,135	(0, 1 dummy variable)
Have job	38,245	10, 11,12,135	(0, 1 dummy variable)
Race:			
White	38,245	10, 11,12,135	(0, 1 dummy variable)
Black African American	38,245	10, 11,12,135	(0, 1 dummy variable)
Other race	38,245	10, 11,12,135	(0, 1 dummy variable)
Marital Status:			
Married	38,245	10, 11,12,135	(0, 1 dummy variable)
Married spouse absent	38,245	10, 11,12,135	(0, 1 dummy variable)
Live with partner	38,245	10, 11,12,135	(0, 1 dummy variable)
Separated	38,245	10, 11,12,135	(0, 1 dummy variable)
Divorced	38,245	10, 11,12,135	(0, 1 dummy variable)
Separated Divorced	38,245	10, 11,12,135	(0, 1 dummy variable)
Widowed	38,245	10, 11,12,135	(0, 1 dummy variable)
Never Married	38,245	10, 11,12,135	(0, 1 dummy variable)
Sociability:			
Participate in religious services	38,245	10, 11,12,135	(0, 1 dummy variable)
Have a good friend	38,245	10, 11,12,135	(0, 1 dummy variable)
Depressed	38,245	10, 11,12,135	(0, 1 dummy variable)
Have Child	38,245	10, 11,12,135	(0, 1 dummy variable)
Total income and wealth:			
Total income in a year	38,245	10, 11,12,135	Total family income in a year
Total wealth(thousand dollar)	38,245	10, 11,12,13	Net value of financial wealth (total wealth net of total debts)
Financial activities:			
Have stocks or mutual funds	38,245	10, 11,12,13	If people have stocks or mutual funds (0, 1 dummy variable)
Have mortgage	38,245	10, 11,12,13	If people have a mortgage (0, 1 dummy variable)
Have checking or saving account	38,245	10, 11,12,13	If people have a checking or saving account (0, 1 dummy variable)
Have bond	38,245	10, 11,12,13	If people have any bonds(0,1 dummy variable)
Have financial advisor	9,561	13	People who trust on financial advisor in their financial activities (0,1 dummy variable)
Degree of Risk averse:			
Risk taker	19,122	12, 13	If people rank themselves 5 or above in variable B132(0, 1 dummy variable)
Internet penetration	38,245	2010, 2012, 2014, 2016	Number of high speed Internet connection per 1000 households in each county. Rank between 0 to 5
Internet use	38,245	10, 11,12,13	If people using the Internet to send or receive email or any other purpose(0,1 dummy variable)
Lightning density	38,245	2010, 2012, 2014,2016	Number of cloud to earth lightning strikes per 1km$\hat{2}$ in a year
Percentage of families with TV in 1950	38,245	1950	Percentage of families who have television

Table 2-A.2: Validity check: Broadband penetration and stock market participation for different levels of speed

	(> 10 Mbps downs and > 1 Mbps ups)				(> 200 kbps in at least one direction)			
For speeds	OLS		IV		OLS		IV	
Stock market participation	0.0264***	0.0333***	0.0475**	0.0450**	0.0357***	0.0470***	0.0582***	0.0513**
	(0.00512)	(0.00725)	(0.0240)	(0.0221)	(0.00604)	(0.00117)	(0.0145)	(0.0252)
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓
Household level controls	✓	✓	✓	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE		✓		✓		✓		✓
N	7,830	7,830	6,631	6,631	7,830	7,830	6,631	6,631

Notes: Table presents the results for the reduced effect of broadband penetration on stock market participation for different speeds. Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth and the number of children. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 2-A.3: Alternative standard errors: Investment in stock market or mutual funds

	Panel A: Standard errors clustered at county level				Panel B: Standard errors accounting for spatial correlation (100km)					
	OLS		IV		OLS				IV	
					Conley (1999)		Colella et al.(2019)		Colella et al.(2019)	
Broadband penetration	0.0264***	0.0333***	0.0475**	0.0450**	0.0264***	0.0333***	0.0264***	0.0333***	0.0475**	0.0450**
	(0.00512)	(0.00725)	(0.0240)	(0.0221)	(0.00580)	(0.00712)	(0.00578)	(0.00637)	(0.0211)	(0.0222)
N	7,830	7,730	6,631	6,631	7,830	7,830	7,830	7,830	6,631	6,631
Individual level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
household level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Survey-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual FE		✓		✓		✓		✓		✓
Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Table presents the results for the effect of broadband penetration on stock market participation, correcting for spatial correlation in standard error. we considered Conley (1999) and Colella et al. (2019) approach for OLS and IV estimates.Individual level controls are Gender, race, age, education, migration, sociability, and retirement status. Household level controls are Household total income, household total wealth and the number of children. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Chapter 3

Roads, Electricity, and Jobs: Evidence of Infrastructure Complementarity in Sub-Saharan Africa

3.1 Introduction

Roads and electricity play central roles in economic development, yet evidence about whether their expansion is a potential driver of job creation is limited and mixed. Moreover, little is known about the complementary effects that such investments may confer in combination or the constraints on the potential benefits of one kind of infrastructure because of the absence or inadequacy of the other type. For example, greater access to electricity can stimulate local production. However, a lack of reliable transport infrastructure can reduce the potential gains due to the lack of market access, especially in more isolated areas. On the other hand, greater connections to road networks can increase market access, allowing for better access to customers but also greater exposure to competition; without access to modern production technology that electricity enables, the competition effect from improved road access may outweigh the benefits of reaching more potential customers, potentially stymieing local economic growth. Empirical evidence on the complementarity between roads and electricity is limited (Moneke (2020)). Most existing studies either analyze the effects of infrastructure investments in isolation, or estimate the aggregate impact of infrastructure in a bundle of many types of infrastructure, calculating the elasticity of output with respect to a synthetic infrastructure index that includes electricity, transport, and telecommunications (Calderon et al. (2015)). Thus, questions remain about whether a big push for investments in roads and electricity represents a powerful means of creating more and better jobs.

This paper investigates how infrastructure investments can be coordinated to support economic development by examining the combined causal effect of electricity and road expansion on employment in sub-Saharan Africa. Our main analysis relies on two sets of geo-referenced household survey data: the Living Standards Measurement Study (LSMS) surveys and Demographic and Health Surveys (DHS). We combine these with an extensive array of data on the geo-location of road expansions (Jedwab and Storeygard (2021)) and electric grids in 27 African countries.

The high costs and potentially large benefits of infrastructure investments mean that the placement of electrification projects and new roads typically correlates with both economic and political characteristics of locations (Blimpo et al. (2013); Burgess and Donaldson (2012); Okoye et al. (2019); Selod and Soumahoro (2019)). To address the potential endogeneity of infrastructure investment, we use two identification strategies: a baseline regression strategy and an instrumental variable (IV) approach. In the baseline model, we account for country and year fixed effects and consider a wide range of individual and local factors as controls.

In the instrumental variable (IV) model, we employ topography data as the main instrument for electrification decisions (Dinkelman (2011)). Additionally, we incorporate the density of lightning strikes, which is a crucial factor contributing to electrical line damage and service disruptions, particularly in Africa (Manacorda and Tesei (2020); Agrawal (2021)).

We propose a new instrumental variable for the proximity to the road network, inspired by post-independence leaders' idea of constructing roads to govern the entire expanse of their countries from the capital city - the center of political power - and to promote ethnic and regional integration and nation building. One of the main goals of African leaders as their countries gained independence was to extend the exercise of their authority over all locations. In the wake of the arbitrary partition of Africa between 1884 and 1885, African leaders also sought to integrate different ethnic groups and regions into their countries (Herbst (2014); Deng (1997)). One way to achieve both policy goals was to construct roads to facilitate the travel of administration officials and military troops from the capital city to other regions of the country and the movement of people and goods between different ethnic homelands. Leveraging this premise for our analysis, we construct hypothetical roads linking ancestral ethnic homelands from the capital of each country to instrument the actual road network. We use a least-cost algorithm, assuming that the slope of the land is the main geographic determinant of road construction costs. The exclusion restriction is likely to be met because these hypothetical roads are conditional only on the geographic costs of construction, not demand-side factors. Moreover, the set of ethnic groups and homelands in each country is exogenously determined, given the arbitrary partition of Africa. However, We conduct several robustness checks to confirm the validity of our choice of instruments.

Estimates suggest positive and significant employment effects of proximity to roads, proximity to electricity grids, and the interaction of the two. For people in locations sufficiently close to an electric grid, bringing the main road 10 km closer increases the probability of being employed by between 2.8 percentage points (based on the LSMS sample) and 20 percentage points (based on the DHS sample). Greater road access also increases the number of hours worked by 1.3 hours per week (LSMS sample). For locations sufficiently close to a main road, bringing an electric grid 10 km closer increases the likelihood of being employed by 7.8 percentage points (LSMS sample) and 19 percentage points (LSMS sample). Greater proximity to an electric grid also raises the number of hours worked by 19.8 hours per week (LSMS sample).

Our estimates also reveal complementarity. We show that bringing an electricity grid closer to a location increases the positive marginal impact of bringing a main road closer, and, in turn, bringing a main road closer to a location increases the positive marginal impact of bringing an electric grid closer. Specifically, we find that bringing an electric grid 10 km closer increases the probability of working by more than 18.9 percentage points in locations with direct road access but by only 12 percentage points in locations more than 10 km away from a main road (DHS sample). Similarly, bringing the main road network 10 km closer increases the probability of working by more than 20 percentage points when that location is near an electric grid but by only 14 percentage points when it is located more than 10 km away from an electric grid. The results suggest that road and electricity expansions play complementary roles in job creation.

We also show that infrastructure investments impact employment structure and spur structural transformation at the local level. That is, better access to road and electricity networks increases employment in skilled occupations and decreases employment in unskilled occupations, especially self-employment in agriculture. Roads and electricity again play complementary roles. Together, they raise the probability of being employed with a skilled job and decrease the probability of having an unskilled job.

Taking advantage of the fine granularity of the available data, we analyze the heterogeneous effects of infrastructure investments between more-developed and less-developed countries, between rural and urban areas, and across gender and age groups. We find that expanding access to roads and electricity, separately and jointly, has a greater positive impact on employment in more-developed African countries than in less-developed countries. Our comparison of rural and urban locations shows that greater proximity to a main road has a greater effect on rural households, increasing skilled employment and reducing unskilled employment. We find that the employment gains due to infrastructure investment accrue to a greater degree for women than for men (DHS sample). Our results also suggest that younger adults and those in middle age benefit more from infrastructure investments.

We conduct several checks to confirm the validity of our instruments. In particular, we examine the extent to which our hypothetical road instrument correlates with early development indicators which might also predict employment. The indicators we consider include the slave trades (Nunn, 2008; Nunn and Wantchekon, 2011), the presence of a city in an ethnic homeland in the precolonial era, and local population density in the early twentieth century. We find none of these variables to be correlated with our instrument for the road. Finally, we check if our instruments for electricity correlate with measures of local economic development. We find no correlation between lightning density and months of the growing season and gross production.

By 2035 Africa's working-age population is expected to grow by 450 million people, an increase of close to 70 percent. Forecasts indicate that the number of people of working age will grow faster than the number of jobs available for them (World Economic Forum, World Bank, and African Development Bank (2017)). Thus, job creation is a pressing priority for African countries. Our findings suggest that investments in roads and electricity are mutually beneficial drivers of employment and structural transformation in Africa. However, the benefits of infrastructure investment could differ significantly across countries, between rural and urban areas, and across gender and age groups.

This paper adds to a growing body of research examining the impact of various infrastructure investments on development outcomes. Recent studies have shown that highways and railroads can have substantial impacts on the allocation of economic activity, land use, local economic specialization, and migration Fajgelbaum and Redding (2014); Cosar and Fajgelbaum (2016); Dave Donaldson (2016); Blankespoor et al. (2017); Berg et al. (2018); Donaldson (2018); Morten and Oliveira (2018); Fajgelbaum and Schaal (2020); Jedwab and Storeygard (2021). We use the same road panel as used by Berg et al. (2018) and Jedwab and Storeygard (2021). In a study conducted by Berg et al. (2018), they found a modest impact of improved market accessibility on local cropland expansion. Additionally, they discovered suggestive evidence of a positive association between improved market accessibility and local GDP growth, indicating the potential stimulation of non-agricultural activities. Study by Jedwab and Storeygard (2021) demonstrate a positive 30-year elasticity of city population in relation to market access. These estimates represent average effects. Recent studies, as surveyed by Berg et al. (2017), have examined the significance of heterogeneous effects. However, most of these studies have focused on investigating spatially heterogeneous impacts. The findings suggest that transportation infrastructure enhances the accessibility of rural regions to cities, potentially resulting in the movement of productive capital and skilled labor from rural areas due to the agglomeration effects of cities. Faber (2014) finds that NTHS highway connections in China have led to a reduction in local GDP growth and industrial output among peripheral counties on the way between targeted metropolitan centers relative to non-connected peripheral counties. A few papers have looked at the heterogeneous impacts across industries. Using historical U.S. data, Chandra and Thompson (2000) find that connections to highways have heterogeneous effects across industries. Ghani et al. (2016) find different impacts from the upgrades to India's Golden Quadrilateral highway network on the organization and efficiency of manufacturing activity across districts, depending on their distance to the Golden Quadrilateral roads. Our paper contributes to this literature by looking at the heterogeneous impacts of roads on outcomes at the household level.

There is a lack of consensus on the benefits of electrification. While many studies find substantial benefits from electrification on income, labor supply and education outcomes

(Dinkelman (2011); Van de Walle et al. (2017); Zhang (2018), others suggest that these benefits may be overstated (Burlig and Preonas (2016); Aklin et al. (2017)). However, a recent paper by Fried and Lagakos (2020) quantifies the long-run, general equilibrium effects of eliminating power outages on firm and aggregate productivity in Nigeria, for which they find much larger estimates than previously estimated in the literature. In line with this paper, our work analyzes the impacts of better access to the national grid for multiple African countries.

In this paper, we investigate how the impacts of infrastructure change when considered in isolation or, instead, when bundled together. A study by Moneke (2020) is one of the few studies to have considered interactions of infrastructure investments. Using data from Ethiopia, he shows that the combination of market access provided by roads and the positive productivity effect of electrification on non-agricultural production is key for the achievement of large effects of infrastructure on economic development. Our paper complements Moneke (2020) by expanding the geographic analysis to 27 countries, and by considering the outcomes at the household rather than district level. To the best of our knowledge, our paper is the first to study the impact of roads and electricity and their interaction at the continental level in Africa and to document significant heterogeneities. We find that isolated and bundled infrastructure effects differ significantly - between lagging and more advanced regions, across gender, and among various age groups. We complement the findings from the literature that lagging and leading regions benefit differently by documenting and comparing these impacts when investments are bundled and when they are isolated. For instance, Mu and van de Walle (2011) find that, in Vietnam, benefits from rural road improvements are greater in poor communities, where levels of initial market development are lower. Moreover, we find that improved infrastructure in rural areas induces a shift away from agricultural and other low-skilled jobs into more skilled occupations; and, by contrast, that improved infrastructure in urban areas promotes all types of employment. These differential effects suggest that the structural transformation brought about by road and electricity expansions is primarily a rural phenomenon. These findings are new.

The remainder of this paper is organized as follows: Section 3.2 describes a simple conceptual framework of infrastructure and job creation and findings from the related literature. Section 3.3 presents data sources. Section 3.4 discusses identification strategies. Sections 3.5 and 3.6 present findings on the average and heterogeneous effects of employment outcomes from infrastructure investment, with a special focus on the contrasts between the effects on urban and rural economies. Section 3.7 represents some robustness check and shows evidence in support of our identification assumptions. Section 3.8concludes.

3.2 Conceptual framework

Investment in infrastructure could influence employment outcomes through a range of different channels and effects. First, infrastructure programs can lead to direct employment on-site for construction, maintenance, and operation. Second, large investments in infrastructure could stimulate short-term growth and indirectly contribute to job creation in related industries (Romer and Bernstein (2009)). Third, and perhaps most importantly, infrastructure could induce employment expansion and enhance the quality of employment by boosting productivity, improving competitiveness, and enhancing market access for both households and firms. Jobs created under this category tend to be longer term and broad based. The rest of this section lists the main economic channels through which access to roads and electricity affects long-term employment, the interactions between both types of investments, and their heterogeneous impacts across population groups and locations. While not exhaustive, this discussion seeks to address the main economic channels of interest to set the stage for the interpretation of the results of our empirical analysis.

3.2.1 Different economic channels

Infrastructure and human capital accumulation Employment is a function of labor supply and demand. Infrastructure affects the quantity and quality of the labor supply through its impact on households' behavior, skills development, and access to the job market. For example, rural electrification increases the productivity of home-production activities, and releases females' time from home- to market-based work (Dinkelman (2011)). Rural electrification is also associated with better health and education outcomes, leading to an improvement in the quality of the labor supply and increased employability of workers. This connection has been explored in studies such as Samad and Zhang (2016) and Samad and Zhang (2018). Better roads are expected to improve households' employment outcomes by reducing the duration of transportation to jobs, allowing workers to access more potential offers, and improving the potential to match workers' skills and available jobs.

Infrastructure and market access Investments in new transport infrastructure may reduce the cost of transportation (Martincus et al. (2017)) and improve market access in both rural and urban areas (Hjort and Poulsen (2019); Aggarwal et al. (2018); Baum-Snow et al. (2016)). Many rural locations in poorer countries lack all-weather roads that connect villages to one another, and/or connect villages to market centers in towns and cities. Isolated villages operate in an environment of near autarky, with product markets too thin for farmers to rely on them to diversify production. Farmers adopt subsistence farming, and produce most of what they need for consumption, instead of specializing in a few crops, and buying the rest from outside (Kebede (2020)). From a similar reasoning at the national scale, Tombe and Zhu (2019) show that the construction of new highways in China increased domestic market integration and had large impacts on population and local GDP.

Infrastructure and productivity Infrastructure could also increase the number and quality of jobs by raising the productivity and competitiveness of businesses and firms. For example, an increase in the adoption and quality of electricity can improve firms' productivity - in turn increasing their net output and demand for labor (Grainger and Zhang (2019); Hjort and Poulsen (2019)). Investments in roads between rural villages and urban markets can reduce the price of inputs to agricultural production, and incentivize the adoption of better technologies that can increase productivity. From work undertaken in rural, northern Tanzania, Aggarwal et al. (2018) estimate that halving travel costs - the approximate effect of paving rural roads - would double the adoption of agriculture-improving technologies, and reduce the adoption-remoteness spatial gradient.

However, the effects might be more complicated when considering labor mobility at both rural and urban levels. For example, Asher and Novosad (2020) shows that a major road construction program in rural India, rather than facilitating the growth of non-farm firms in treated villages, enabled workers to access external labor markets, and supported labor reallocation to (non-farm) wage labor. Moreover, the average effects of infrastructure investments have been found to be positive in some cases and negative in others. Reduced-form estimates of the impacts of new highways in China by Baum-Snow et al. (2016) indicate that the expansions of regional highway networks had negative average effects on local populations and no significant effects on local GDP. A recent strand of the literature has shown that average impacts can be misleading with some groups or regions benefiting more than others, suggesting that economic impacts can differ significantly across locations and groups (Baum-Snow et al. (2016); Lall and Lebrand (2020)).

3.2.2 The heterogeneous impacts of infrastructure investments

Heterogeneities across gender and age groups The impacts of infrastructure provision may differ across genders. Electrification has been shown to disproportionally benefit women and girls. In households with electricity, women spend less time on household chores, and they are more likely to participate in income-generating activities; girls have higher educational attainment (see Samad and Zhang (2017) and Samad and Zhang (2018)). On the other hand, gender-based constraints and preferences could limit opportunities and benefits to women from investment in infrastructure. For example, women may play a smaller part in labor-intensive forms of infrastructure construction (which is often linked to employment creation) than their male counterparts. Similarly, the impacts on employment differ across age groups. When investments in infrastructure boost new economic activities at the expense of old activities, older workers are more likely to lose their jobs. Changing sectors may be more difficult for older workers than those in their prime. The employment effects of an investment in infrastructure could also differ according to workers' characteristics, such as skill level. Lowering transportation costs and expanding the supply of electricity may speed up the structural change of economies, with jobs moving from low-productivity agricultural sectors to manufacturing or service sectors. The relative demand for skilled manufacturing workers may therefore increase, while the relative demand for unskilled workers in the agriculture sector may decrease.

Heterogeneities across locations Economic impacts of infrastructure investments vary across locations. Investing in rural areas, in nationwide projects, or in urban projects will activate different channels to generate jobs. Rural roads can trigger the structural transformation, with workers moving away from subsistence agriculture to services. National roads may support jobs in manufacturing and export-led industries. The potential of roads to match skills and jobs might differ, with different ramifications for isolated areas, the periphery of cities, and urban contexts. For example, improving access to jobs and enhancing skill matching in cities allow the density of the population to create positive spillovers from agglomeration large enough to surpass the negative spillovers from congestion. Similarly, the gains from investing in modern energy will differ for firms in large cities, and for small businesses in villages and rural areas. While the type of investment matters, its location will also affect the types of impacts one can expect to occur. The impacts of road construction have been shown to be heterogeneous across locations. Baum-Snow et al. (2016) provide reduced-form estimates of the impacts of new highways in China. While the average impact is negative, they show that it results from a large heterogeneity in the effects of highways on prefectures. Regional highways promote the concentration of both output and population into regional prefectures, at the expense of other prefectures. Since there are more small prefectures than large ones, the average is negative.

3.2.3 The complementarities between road investments and electrification

Investments in roads or electrification alone may have limited or even adverse impacts in some contexts, for some locations, and for some groups. On the one hand, electricity can stimulate local production. However, the lack of reliable transport infrastructure will reduce the potential gains, especially in more isolated areas. Electrification alone mostly increases productivity in remote locations that have high transport costs and tend to remain shielded from import competition (Behrens et al. (2006)). While some positive effects driven by local demand are predicted from, say, greater access to electricity, these electrified locations may miss out on other regions' increased import demand for their newly electrified goods simply because they lack the transportation connection.

On the other hand, connecting isolated locations may increase import competition and have adverse impacts on local economic activities and jobs. Using the insight from a spatial quantitative model, Moneke (2020) shows that in partial equilibrium, as previously autarkic regions gain access to markets, the pre-existing employment in the manufacturing sector suddenly competes with manufacturing-sector varieties from larger, more productive agglomerations. Unless the initial manufacturing-sector productivity is high, the sectoral-employment share of the manufacturing sector in the location newly connected by roads would be expected to fall. However, as productivity in previously isolated locations improves in the wake of new road connections, the roll-out of electrification can lead some manufacturing sectors to become profitable for exports, and to such a degree that the manufacturing employment share may actually rise. In the context of Ethiopia, reduced-form analysis by Moneke (2020) shows that road access alone causes manufacturing employment to drop largely in skilled activities, and, by contrast, causes service employment to increase, largely in informal, small-scale retail jobs.

The enabling effects of an investment in infrastructure on job creation may depend on the interactions of investments. Whether "big push," combined investments deliver higher benefits than isolated investments of only roads or only electrification remains an open empirical question. One of the few attempts to examine the issue is the work in Ethiopia by Moneke (2020), who finds that only the interactions of infrastructure investments - not isolated investments - gave rise to large effects on economic development.

Thus, this paper aims to complement existing empirical studies and to fill a research gap, by looking at both the average and heterogeneous impacts of investment in infrastructure, and by examining the complementary effects between two types of investments, roads and electricity.

3.3 Data

To analyze the effects of infrastructure on employment in sub-Saharan Africa, we use two geo-referenced, household surveys: the Living Standards Measurement Study (LSMS) surveys conducted between 2005 and 2019, and the Demographic and Health Surveys (DHS) conducted between 2000 and 2018. Both surveys are merged with spatial information on roads and electricity grids.

3.3.1 Demographic and Health Surveys (DHS)

Demographic and Health Surveys are repeated, cross-section surveys that have been conducted in most developing and middle-income countries since the mid-1980s. These surveys are representative at the national and sub-national levels. Information on demographic and socioeconomic status (e.g., age, education, occupation) is collected for all household members. In addition, selected women and men provide information on a wide range of other variables, including matters concerning health status, fertility history, migration, and children. The DHS Program routinely collects geo-referenced data (longitude and latitude) in selected surveyed countries, making it possible to combine these surveys with a wide range of external data and information. DHS has been used in recent studies to analyze the economic and social impacts of infrastructure in developing countries (Okoye et al. (2019); Hjort and Poulsen (2019); Moneke (2020); Canning et al. (2020b); Herrera Dappe and Lebrand (2021); Lebrand (2022)).

We use data from 64 surveys that cover 26 countries in sub-Saharan Africa. The combined sample includes 840,125 women and 373,745 men.[1] Table 3-1 summarizes the characteristics of the individuals in the sample.

We categorize employed individuals as being either high-skilled or low-skilled workers. We also distinguish agricultural workers in parts of our analysis. Employed individuals are considered to be high-skilled workers if, at the time of the survey, they worked in any of the professional, clerical, skilled manual labor, sales, services, or agriculture (employee) occupations and low-skilled employees working in any of the domestic, unskilled manual labor, or agriculture (self-employed) occupations. Among the respondents in our data, 69.2 percent reported being employed, and 30.97 percent reported that they were not employed. Skilled workers constitute 32 percent of the sample. Among them 38 percent have sales occupations, 18 percent have manual jobs, 16 percent are agricultural employees, 12 percent

[1]The number of observations is greater for women than men because women are over-sampled in the DHS.

work in service-sector, professional, technical or managerial occupations, and 2 percent work in clerical occupations. Among unskilled workers, 80 percent reported being self-employed in agriculture, and the remaining 20 percent were either in domestic work or as unskilled manual laborers. Among respondents, 63 percent resided in rural areas, and 39 percent resided in urban areas. The age of survey respondents is as follows: 25 or younger (39 percent), 25 to 50 (57 percent), over 50 (0.2 percent), and the mean age of the respondents is 32.34 years old.

To look at the differential effects of an investment in infrastructure across countries at different stages of development, we group the survey data into two development categories. Using PPP-adjusted GDP per capita, we classify countries as less developed or more developed. A country is less developed if its PPP-adjusted GDP per capita falls below 2017 USD 3,400. In the LSMS, 97.37 percent are categorized as less developed. In the DHS, 56 percent of the countries are categorized as less developed. An average household is located 9.97 km away from a main road and 26.19 km away from an electric grid.

3.3.2 Living Standards Measurement Study (LSMS)

LSMS surveys are conducted by the World Bank. The LSMS project collaborates closely with the national statistics offices of its eight affiliated countries in Sub-Saharan Africa. The project's primary focus is to implement comprehensive, nationwide representative panel household surveys that cover multiple subject areas, with a particular emphasis on agriculture. The overarching goal is to foster innovation and productivity in statistical research by examining the interconnections between agriculture, socioeconomic conditions, non-agricultural activities, and poverty reduction.

The decision to conduct data collection is made on a country-specific basis, taking into account data requirements and the availability of funding. The sampling methodology employed follows a stratified two-stage approach, utilizing a two-stage probability sample. In the initial stage, enumeration areas (EAs) are selected with probabilities proportional to the total number of households within each EA. Subsequently, households are chosen within each selected EA to establish a cohort that can be followed and reassessed over time on various aspects such as behavior, welfare, income distribution, consumption patterns, employment, health, education, and financial decision-making.

In the second stage, a subset of the primary household sample is randomly selected within each EA using equal probabilities. Over subsequent survey years, the project conducts follow-up assessments of households residing within the same EA. However, due to household mobility, some households may have exited the survey, while others may have entered. The process of drawing households within the EA is performed after a listing operation, which serves to update the selection of households and ensure accurate representation.

The LSMS project's approach combines rigorous sampling techniques and panel survey methodology to gather robust data that can support evidence-based research and policymaking in the field of agriculture and its broader socioeconomic implications.

We use 24 LSMS surveys from 6 sub-Saharan African countries: Ethiopia, Malawi, Niger, Nigeria, Tanzania, and Uganda. All the surveys and their corresponding years are listed in Table 3-A.1. The collection of information and follow-up across subsequent surveys vary across countries. In our sample, 51 percent of respondents are women, and 86 percent of respondents reside in rural areas. Roughly 58.8 percent of individuals who responded to the survey were working at that time. The mean working hours during the last 7 days was 2.193 hours. Among these individuals, around 50 percent were working in the agricultural sector. Over 50 percent of individuals have at least a secondary-level education.[2] The average age of a respondent is 36.12 years old. Around 78 percent of sampled individuals are between 20 and 45; around 22 percent of individuals are between 46 and 66.

We primarily use data on respondents' occupations. Following Hjort and Poulsen (2019), we categorize employed individuals as being either high-skilled or low-skilled workers. Also, given the importance of agriculture in African economies, agricultural employment is distinguished from other types of employment in part of the analysis. Respondents were classified as low-skilled workers if they worked in the following occupation types: farming, fishing, and domestic employment. At the time of the survey, among employed individuals, 63 percent are classified as high-skilled workers, and 49 percent are classified as low-skilled workers. Farmers represented 50 percent of workers.

Using the geo-referenced information (latitude and longitude) associated with these surveys, we combine them with the roads and electricity datasets. An average household is located 15.43 km away from a main road and 8.62 km away from an electric grid.

3.3.3 Roads

We combine the geo-referenced households from the DHS and LSMS with the information on road network expansions as used by Berg et al. (2018) and Jedwab and Storeygard (2021).[3] The development of road infrastructure is tracked over 40 years, from 1971 to 2014. Data on layout and road type are included, relying on layouts from Nelson and Deichmann (2004)

[2]Details are displayed in Table 3-1.

[3]The roads digitization effort of Jedwab and Storeygard (2021) benefited from funding under the DFID/World Bank's Strategic Research Program on Transport Policies for Sustainable and Inclusive Growth.

and completed with information extracted from Michelin maps. The road-layout information used by Nelson and Deichmann (2004) was primarily constructed using information from the U.S. Digital Chart of the World database, originally constructed by the U.S. Defense Mapping Agency. Road surface information was obtained using 64 Michelin road maps for the northeast, northwest, and central-south regions of mainland sub-Saharan Africa. Roads are divided into the following categories: earth/dirt, paved, improved, or highway. We combine the last three categories into one "main road" category. Michelin also collects statistics on road development projects, and its maps are meant to specify the year during which a specific road was made operational. However, Michelin maps for some states and time periods do not exist. Overall, the road network in Africa has steadily expanded from 239,018 km in 1971 to 334,391 km in 2014. Figure 3-1 depicts the evolution of main roads from 1971 to 2012. Overall, the mean distance of households from the main roads is 9.978 km in the DHS and 15.436 km in LSMS surveys.

3.3.4 Electricity

Information on the expansion of the electricity grids is obtained from World Bank/Energy Sector Management Assistance Program (ESMAP) data sources. Arderne et al. (2020) present the first composite map of the global power grid using publicly available open data. The map was generated using data obtained through Gridfinder and OpenStreetMap data. Gridfinder is an open-source tool for predicting the location of electricity network lines using satellite imagery of nighttime light. Combined with OpenStreetMap data, multiple filtering algorithms are applied to nighttime light imagery to identify locations most likely to produce light from electricity. These light sources (target locations) are then connected to known electricity networks through a least-cost routing algorithm following roads and known distribution lines (adopted from OpenStreetMap). This results in connected networks at two voltage levels: high voltage and medium voltage. The data we use is for the year 2010. Time variation in connectivity to electricity within a country, therefore, stems from the use of multiple rounds of surveys for each country and survey samples that vary across years. Hjort and Poulsen (2019) follow a similar design. The OpenStreetMap data show that 97 percent of the global population lives within 10 km of electricity lines. Globally, an estimated 2 percent of the population does not live within 10 km of electricity lines; most of this population lives in sub-Saharan Africa. In our data set, the mean distance of households to the closest electricity grid is 25.79 km for the DHS survey and 7.90 km for the LSMS survey.

Figures 3-2 and 3-3 show how employment relates to proximity to a main road and proximity to an electricity grid. To generate these figures, we compute the probability of employment for each given distance to an infrastructure point, where distance is measured in bins of 10 km. We then fit the prediction of employment on distance. As the distance to main roads increases, the probability of having jobs decreases. This is true for both LSMS and DHS samples. In the LSMS sample, for example, those living within 10 km of a main road were 2.4 times more likely to be at work than those living within 90 to 100 km of a main road. A similar pattern is observed in the DHS sample in Figure 3-3. The patterns are similar with regard to the relationship between employment and distance to an electricity grid. We also note that the proximity to roads and electricity positively affects the probability of working in a high-skilled occupation and negatively affects the probability of working in a low-skilled occupation.

3.4 Identification strategies

We use two main identification strategies to analyze the causal effects of access to main roads, access to electricity, and the interaction of the two on job creation in sub-Saharan Africa. The first and baseline strategy is ordinary least squares (OLS) with individual-level controls and country and year fixed effects. The second strategy develops an instrumental variables approach.

Dependent variables. We analyze five main dependent variables. The first measures whether an individual (in a specific location and time period) is employed. The second accounts for the number of hours worked during the last seven days. The third variable measures whether an employed individual is a farmer. The fourth variable measures whether a worker is unskilled, and the fifth focuses on higher-skilled employment categories. We follow Hjort and Poulsen (2019) to create an intermediate employment category between high-skilled and unskilled.

Independent variables. We analyze the effects of three main independent variables - proximity to a main road, proximity to an electricity grid and the interaction of the two - on the dependent variables. Proximity to a given infrastructure is measured by taking the negative value of the distance between the centroid of an individual's cluster and the closest location of the corresponding infrastructure.

3.4.1 Baseline strategy: OLS with fixed effects

For our baseline strategy, we estimate the following equation using OLS:

$$Emp_{ict} = \alpha_0 + \alpha_1 Road_{ict} + \alpha_2 Elect_{ict} + \beta(Road_{ict} \times Elect_{ict}) + X'_{ict}\delta + \mu_c + \theta_t + \epsilon_{ict} \quad \text{(3-D.1)}$$

Where Emp_{ijct} is the outcome variable, measuring whether individual i in the country c and at year t is employed. In regressions using skilled/unskilled worker as the dependent variable, Emp_{ict} measures whether individual i in the country c and year t is a skilled or unskilled worker. $Road_{ict}$ measures proximity of individual i in country c at time t to the road networks, and $Elect_{ict}$ measures proximity of individual i in country c at year t to the electricity grids. Variable $Road_{ict} \times Elect_{ict}$ measures the interaction between proximity to the road and proximity to the electricity grid simultaneously. [4]

In our main specification, the parameters α_1 measure the marginal effect on employment of moving 1 km closer to the road networks in locations within zero km of an electricity grid, and α_2 measure the marginal effect on employment of moving 1 km closer to the electricity grids in locations within zero km to a main road.[5] The parameter β measures the incremental effect of moving closer to the road networks and the electricity grids. If all these parameters are positive, then roads and electricity are complements in the process of job creation. If, on the contrary, β is negative while both α_1 and α_2 are positive, then the two types of infrastructure are substitutes. Being complements (substitutes) means that one type of infrastructure is more (less) useful when the other type is supplied.

Our identification strategy exploits spatial and temporal variation in the expansion of the road networks and mostly spatial variation in the expansion of the electricity grids.[6] The term μ_c captures country fixed effects; and θ_t captures survey year fixed effects. Controlling for year-fixed effects accounts for all the time-variant factors common to all countries, which may affect infrastructure and job creation. X'_{ict} is a vector of individual-level controls, including age, gender, type of place of residence, education, average temperature, and distance to the sea. The identifying assumption is that the error term, ϵ_{ict}, is uncorrelated with each of the variables measuring proximity to infrastructure given all the aforementioned controls.

[4] Electricity grids are often located along main roads, which can make the identification of having access to the grid alone difficult (Moneke (2020)). In this paper, we keep all three variables, proximity to the road, proximity to the electricity grid and their interaction term of the following reasons. First, we consider the distance to the closest road, which might be different from the distance to the road along which the closest electricity grid point can be found. This analysis differs from Moneke (2020), which uses binary access variables and finds that all districts with access to the electric grid have a main road, which makes identifying access to the grid alone impossible. Second, the correlations between our three variables are positive but not very high. For example, the correlation between proximity to the road and proximity to the electricity grid is 0.38 in the LSMS sample and 0.34 in the DHS sample.

[5] This interpretation stems from the fact that each variable that measures proximity to a given infrastructure takes only "negative" values, with greater values implying greater access; the highest access is therefore achieved when the value of the variable is zero.

[6] We follow a research design similar to that of Hjort and Poulsen (2019), in which temporal variation in connectivity to the electricity grid comes from the fact that survey samples vary across years in the DHS and the LSMS for each country.

3.4.2 Instrumental variables approach

For our second identification strategy, we use an instrumental variables approach. In fact, inferring causation from the OLS regression requires that proximity to the road networks and the electricity grid be uncorrelated with the error term of equation (3-D.1) given individuals' characteristics and the country and year fixed effects. This would not indeed be the case if, within each country in a given year, roads and electricity grids were randomly assigned. However, owing to their high costs, road construction and electrification are likely to be determined by the location of mineral resources, local labor market conditions, and trade demand. In the African context, the placement of such infrastructure might also be determined by colonial history and other political factors (Okoye et al. (2019)). It is, therefore, reasonable to assume that locations that receive these infrastructure investments have specific characteristics that determine employment and occupational sectors. To the extent that these local characteristics are time-invariant, our baseline strategy, which controls for country fixed effects, accounts for them. However, if some of these characteristics change over time, this can lead to an endogeneity problem that is unlikely to be completely solved by controlling country and year fixed effects. To account for this possibility, we instrument for proximity to roads using a new instrumental variable and instrument for proximity to electric grids using existing instrumental variables.

Instrumental variable for proximity to road

We generate a new instrumental variable for proximity to the road networks based on the rationale that post-independence African leaders promoted the construction of transportation infrastructure to allow them to govern the entirety of their respective territories from the capital city - the center of political institutions - and to bolster national integration. Starting from the colonial era, the construction of transportation infrastructure, consisting mainly of roads and railways, has primarily been motivated by economic, social, and administrative (political or military) reasons in Africa (Taaffe et al. (1963), Okoye et al. (2019), and several references therein). Colonial powers developed transportation infrastructure to move agricultural products and mineral resources from the countries' interior to ports of exports (Okoye et al. (2019)). In each country, such infrastructure also served an administrative purpose; better roads and railroads made it possible for colonial officials to travel to most parts of the country. Though the same reasons motivated the construction of transportation networks in the post-independence period, country leaders followed a different strategy in expanding these networks. They instead built transportation networks in a way that would allow them to extend the exercise of their authority from the "political center" of the country to their entire territory (Herbst (2014)), including remote areas and all ethnic homelands (Muller-Crepon et al. (2020)). Indeed, the arbitrary partition of Africa by European leaders at the Berlin Conference in 1884 and 1885 endowed the countries that emerged from this experiment with different levels of ethnic heterogeneity. In most of these countries, independence leaders faced the daunting task of governing a multitude of disparate ethnic groups and of unifying and integrating them (Deng (1997)). These leaders used a number of different strategies to promote nation-building and national unity. These strategies included promoting the use of one national language, outlawing political parties organized around ethnic groups, and distributing ministerial posts and other official positions among different ethnic groups to foster national pride (Deng (1997)). Interestingly, building infrastructure such as roads was also a strategy for national integration used in all African countries.[7]

Herbst (2014) argues that the difficulty African political leaders faced in governing countries divided into multiple ethnic groups, coupled with the dissociation of ownership and control of land, motivated them to develop infrastructure that would extend their authority as far as possible (see also Herbst (2000)). The poor or absent state of road networks in most countries in sub-Saharan Africa leaves a majority of rural people cut off from the centers of political power. Herbst (2000) argues that African governments have been building roads that connect political centers to previously inaccessible areas to expand their power.

Work by Muller-Crepon et al. (2020) relies on a theory that the accessibility of ethnic homelands by public authorities is a prerequisite for political leaders to extend their domination over remote areas (Tollefsen and Buhaug (2015); Raleigh and Hegre (2009); Buhaug and Rod (2006); Herbst (2000); Fearon and Laitin (2003)). Indeed Muller-Crepon et al. (2020) argue that political instability is predominant in areas where political power has very little control over the inhabitants; more specifically, they establish that the ethnic groups' homelands are difficult to reach from the national capital are also sites of rebellion and political conflicts. The authors use the map of road networks in African countries to measure access to ethnic homelands from national capitals and the interconnectivity of ethnic groups. Estimating the effects of these two variables on the occurrence of political conflicts, they show that political conflicts reoccur in areas where there is a weak physical presence of the state. To demonstrate the robustness of this finding, they construct a theoretical network composed of hypothetical road segments that ensure the lowest travel time between the busiest ethnic groups; using these hypothetical networks, they construct instrumental variables

[7] Building infrastructure as a means to integrate different ethnic groups is not a phenomenon unique to African countries. In their study on the relationship between ethnicity and economic development in China (Han and Paik (2017)) argue, "The state has pursued provisions of the necessary infrastructure such as roads, railways, and airports as means to promote local economic growth and, in turn, integrate ethnic minorities with the Han majority" (p. 33).

estimating what the accessibility of ethnic homelands from political centers as well as the inter-connectivity of ethnic groups would be if these networks existed.

Their results are robust to the use of these instruments, thus confirming the twofold role of road networks in extending the control of political powers and reducing political instability; political authority and stability are reinforced in ethnic groups that are accessible from the political capital.

We rely on this literature, highlighting the motivation of political leaders to build roads to extend their capacity to govern to construct our instrumental variable for proximity to the road networks. In so doing, we follow a large literature that shows that non-economic factors play determining roles in the propensity of African governments to invest in the construction of public infrastructure (Bonfatti et al. (2019); Jedwab and Storeygard (2021); Burgess et al. (2015); Herbst (2014)). In constructing our instrumental variable, we assume that African leaders would build roads to facilitate territorial administration by easing the movement of administration officials and military troops from the capital power center to the rest of the country and from one location to another (Herbst (2014); Muller-Crepon et al. (2020)). Following these assumptions, we use the map of African ethnolinguistic divisions produced by the American anthropologist George Peter Murdoch in 1959 (Murdock (1959)) to identify the ethnic homelands of each country in our sample. These ethnic homelands are regarded as being exogenous. Using the least-cost path algorithm, we "construct" hypothetical roads that link the centroids of different ethnic homelands to each country's capital city. We proceed under the assumption that the social planner's goal would be to minimize the construction costs of these roads if they were to be built. We assume that slope of land is the main geographic determinant of the construction cost. For each country, this hypothetical road network, therefore, intends to show how such a network would evolve to connect different ethnic homelands with the capital city if only geographic costs of construction were considered and all demand-side factors were ignored.

We use proximity to these hypothetical road networks as an instrument for proximity to the existing road networks. For this variable to be a valid instrumental variable, the hypothetical road networks should be close enough to the actual networks, and these hypothetical networks should be exogenous with respect to job creation. The two conditions are likely to be satisfied in our analysis. For example, figure 3-4 compares the hypothetical road networks to a few countries' 2014 actual road networks. As this map shows, these two networks are close enough to satisfy this condition. Moreover, the first and fourth columns of table (3-2), containing the first-stage diagnosis, clearly show a positive and statistically significant correlation between proximity to a main road and proximity to our hypothetical road (using both the DHS and the LSMS samples).

First-stage results are presented in Table 3-2 for both samples and show a strong and statistically significant relationship between instrumental variables and endogenous regressors. Given that we have multiple endogenous regressors, Sanderson and Windmeijer (2016) provide the most relevant weak instrument F-test statistic. Following the authors, we report in table 3-2 Cragg-Donald test statistics along the Sanderson-Windmeijer statistic for weak instrument and over-identification test, indicate non-weak and valid instruments. It follows that our first stage is strong.

Instrumental variables for electricity

We exploit terrain topography and lightning intensity to generate two instruments for electricity. Using geographical instruments such as weather variations, terrain roughness, or land elevation is common in existing empirical work. Indeed, the random nature of these geographical instruments gives them the same properties as those of exogenous variables. For example, previous research has used exogenous changes in rainfall as an instrument for economic growth (Edward Miguel and Sergenti (2004); Miguel (2005); Paxson (1992)). Other research has relied on terrain elevation as a proxy for the cost of constructing electric grids (Amorim et al. (2018); Durante (2009); Dinkelman (2011); Ostrom (1990)).

To construct the instrument for access to electricity grids, we exploit the fact that lightning is a significant contributor to power disruptions in most sub-Saharan African countries (Alvehag and Soder (2010); Minnaar et al. (2012)), which is strictly exogenous. For example, in South Africa, lightning damage accounts for about 65 percent of all over-voltage damage to electrical distribution networks and is believed to account for one-third of all outages. More than 50% of power outages on transmission lines in Eswatini are attributed to lightning. In Africa, every year, lightning strikes and thunderstorms damage circuits and knock out electric power lines, crippling coverage (Minnaar et al. (2012)). The recovery time depends on the extent of damage; in some cases changing the route of electricity lines is necessary. The recovery period is generally long (Karagiannis et al. (2017)) and maintenance is costly. Hence, frequent lightning strikes in a location negatively affect the electric-energy supply and household access to electricity by damaging electricity grids. Following Andersen et al. (2012, 2011); Andersen and Dalgaard (2013); Agrawal (2021), we use lightning density as an exogenous determinant of power disturbances and disruption to access to electricity that can affect the individuals' location "proximity" to the electricity grids. Note that while some of these papers have used lightning intensity as an instrument for the penetration of the Internet (Agrawal (2021)), the main argument in those papers is that lightning strikes damage electrical infrastructure useful to the functioning of the Internet.

In addition, considering that topography can be viewed as a geographic obstacle that makes infrastructure construction more expensive (Ostrom (1990); Durante (2009); Amorim et al. (2018); Dinkelman (2011)), we add terrain elevation to the exogenous variables that we consider as a determinant of the proximity to electrification grids. Therefore, in our empirical model, the proximity to the electrification grids is partly explained by the topography of the land on which the households are located and the lightning density in that area defined by the mean of lightning strikes per square km each year.

We use satellite data generated by the National Aeronautics and Space Administration (NASA) on lightning strike density and aggregated by Manacorda and Tesei (2020); Cecil et al. (2014). They compute average lightning strike density between 1995 and 2015 in $0.5° \times 0.5°$ for Africa. In the DHS, we matched the closest location mean lightning density for a specific year with an individual's location. As reported in Manacorda and Tesei (2020), Africa has the highest lightning density on earth, with an average of 17.3 lightning strikes per square km per year. The world average is 2.9 lightning strikes per square km per year. The Democratic Republic of the Congo has the highest annual lightning strikes globally. Importantly, as figure 3-5 shows, lightning density varies greatly within African countries. implying that this instrument can generate meaningful variation in electricity access. Data on land elevation are provided in the geo-referenced component of the DHS datasets.

Columns 2 and 5 of Table 3-2 confirm the significant and negative relationships between the proximity to electrification grids on the one hand, and both the elevation of the land and lightning on the other. These negative correlations are found in both the LSMS and DHS samples. In addition, the high values of Cragg-Donald Wald F-statistics and the Windmeijer F-statistics show that our instruments are strong.

Given that proximity to the road networks and the electricity, grids are two potentially endogenous variables; their interaction is likely to be endogenous. To instrument this interaction term, we use the interaction term between proximity to our hypothetical road networks (the instrumental variable used for proximity to the road networks) and terrain elevation (one of the instrumental variables used for proximity to the electricity grids). Column 3 of table 3-2 shows a positive correlation between the instrument and the outcome in both the LSMS and DHS samples. Here the high values of the Wald and Windmeijer statistics also demonstrate the strength of these instruments.

Instrumental variable estimation strategy

To estimate the causal effects of infrastructure on employment, we use a two-stage least squares (2SLS) model. In the first stage, we regress the three main endogenous regressors (proximity to roads, proximity to electricity, and the interaction of both variables) on our instrumental variables, controlling for a wide range of individual and geographic characteristics and for country and year fixed effects. In the second stage, each of our outcome variables is regressed on the predicted values of the endogenous regressors obtained from the first stage, controlling for all the variables controlled in the first stage. Note that these two stages are executed simultaneously (and not separately). More precisely, we estimate the following equations:

$$\begin{aligned}\widehat{Road}_{ict} = {}& \alpha_0 + \alpha_1 HypoRoad_{ict} + \alpha_2 FlashDensity_{ict} + \alpha_3 Elevation_{ict} + \alpha_4 Elevation_{ict} \\ & \times HypoRoad_{ict} + X'_{ict}\delta + \mu_c + \theta_t + \epsilon_{ict}\end{aligned} \tag{3-D.2}$$

$$\begin{aligned}\widehat{Elect}_{ict} = {}& \beta_0 + \beta_1 HypoRoad_{ict} + \beta_2 FlashDensity_{ict} + \beta_3 Elevation_{ict} + \beta_4 Elevation_{ict} \\ & \times HypoRoad_{ict} + \mathrm{X}'_{ict}\pi + \mu_c + \theta_t + \epsilon_{ict}\end{aligned} \tag{3-D.3}$$

$$\begin{aligned}Road\widehat{\times Elect}_{ict} = {}& \lambda_0 + \lambda_1 HypoRoad_{ict} + \lambda_2 FlashDensity_{ict} + \lambda_3 Elevation_{ict} \\ & + \lambda_4 Elevation_{ict} \times HypoRoad_{ict} + X'_{ict}\kappa + \mu_c + \theta_t + \epsilon_{ict}\end{aligned} \tag{3-D.4}$$

$$Emp_{ict} = \gamma_0 + \gamma_1 \widehat{Road}_{ict} + \gamma_2 \widehat{Elect}_{ict} + \gamma_3 Road_{ict}\widehat{\times Elect}_{ict} + X'_{ict}\tau + \mu_c + \theta_t + \epsilon_{ict} \tag{3-D.5}$$

In equations 3-D.2, 3-D.3 and 3-D.4, $HypoRoad_{ict}$, $FlashDensity_{ict}$, $Elevation_{ict}$, and $Elevation_{ict} \times HypoRoad_{ict}$ are respectively proximity to the hypothetical roads (also measured as the "negative" of the distance between the centroid of an individual's cluster to the closest hypothetical road); the density of lightning strikes in the closest location of individual i for which we have data; terrain elevation; and the interaction term between proximity to the hypothetical roads and terrain elevation for individual i, in the country c, in survey year t; X'_{ict} is a vector of individual-level controls including age, gender, place of residence, education, and average temperature and distance to the sea. μ_c captures country fixed effect; and θ_t captures survey year fixed effects. We follow a similar approach used in Moneke (2020) in controlling for local characteristics but we add proximity to the sea border as an important variable affecting the job availability in each location. Table 3-2 presents the first-stage regressions.

3.5 Average treatment effects of roads and electricity

3.5.1 Employment

We present the estimates of the "average" treatment effects of proximity to roads and electricity grids and their interaction on employment outcomes in table 3-3 in the Appendix. The first two columns display the estimated effects on work status at the survey for the LSMS and DHS samples, respectively. The third column shows estimated effects on the number of hours worked during the last seven days (derived from the LSMS sample only because information on the number of hours worked is not collected in the DHS).

These estimates show positive and highly complementary effects of roads and electricity on employment outcomes. Columns 1 and 2 show that for locations close enough to an electric grid bringing road networks 10 km closer marginally increases the probability of working by 20.4 percentage points in the DHS sample and by only 2.7 percentage points in the LSMS sample.[8] Similarly, for locations close enough to the road networks, bringing an electric grid 10 km closer marginally increases the probability of working by 18.9 percentage points in the DHS sample and 7.8 percentage points in the LSMS sample. The estimates obtained from the two samples differ as both surveys cover different sets of countries. Our preferred estimates come from the DHS sample, which covers more countries.

The joint, marginal effect of these two types of infrastructure is positive, indicating that roads and electricity play complementary roles in job creation in sub-Saharan Africa. In other words, there are significant employment gains that arise from providing both types of infrastructure simultaneously. The estimates of table 3-3 in the Appendix imply that being closer to a main road increases the positive effect of being closer to an electricity grid on the probability of being employed. These effects are statistically significant. To put these estimates in perspective, using the DHS sample, figure 3-6 shows that the impact of increasing proximity to the electric grid grows for households that are closer to the road, and vice versa - the impact of increasing proximity to roads grows for households that have greater proximity to the electric grid. Being 10 km closer to the electric grid increases the probability that an individual is working by more than 18 percentage points when the individual's household is located next to a main road but only by 12.4 percentage points when located 10 km away from the main road network. Similarly, being 10 km closer to a main road network increases the probability that an individual is working by more than 20.4 percentage points when that

[8] Note that the coefficient on the variable "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.

individual's household is located next to the electric grid but by only 13.9 percentage points when the household is located 10 km away from the electric grid.

The effects of roads and electricity on the number of hours worked in the last seven days are consistent with their effects on employment. The estimates presented in Column 3 of table 3-3 in the Appendix indicate that in areas close enough to an electric grid, bringing a main road 10 km closer marginally increases the number of hours worked by nearly 1.33 hours a week, but this effect is statistically significant at 90 percent confidence interval. Unlike proximity to roads, proximity to the electric grid has a much larger effect. In areas with high access to a main road, bringing the electric grid 10 km closer increases the number of hours worked by 19.8 hours a week. Importantly, it is more beneficial to provide both types of infrastructure simultaneously, as indicated by the positive coefficient on their interaction term. Indeed, bringing a main road 10 km closer increases the positive effect of bringing the electric grid 10 km closer by more than 3.8 hours of work every week. Using the LSMS sample, figure 3-7 shows that the impact of increasing proximity to the electric grid on the number of hours worked becomes higher for households that are closer to the road and vice versa (the impact of increasing proximity to the roads also boosts the number of hours work for households that have greater proximity to the electric grid). Bringing an electric grid 10 km closer increases the number of hours worked by 19.8 hours for households located next to a main road and 16 hours when households are located 10 km away from the main road network. Similarly, bringing the main road network 10 km closer increases the number of hours worked by 1.3 hours when households are located next to the road, but such proximity actually reduces the number of hours when the household is located 10 km away from the electric grid. That is, increasing the proximity to the road only has a positive impact on workers located next to the electric grid.

We note some consistency between the IV estimates and the OLS results reported in table 3-A.3. In all cases but one, the IV coefficients have the same sign as the corresponding OLS coefficients; the only reversal of sign is the negative effect of electricity on the number of hours worked. Unsurprisingly, the OLS coefficients are smaller, indicating a downward bias.

Summarizing the findings, we note that supplying roads and electricity leads to significant employment and productivity gains in sub-Saharan African countries. The effect of electrification tends to be larger than that of roads. When supplied simultaneously, these two types of infrastructure reinforce each other, resulting in greater efficiency. Indeed, they are highly complementary in the sense that supplying one significantly raises the marginal effect of the other. This complementarity is robust across the two outcomes that we have analyzed and across the two data sources. We next examine heterogeneity in the effects of infrastructure on the type of employment and occupational sectors.

3.5.2 Skilled versus unskilled employment and occupational sectors

We analyze the separate and joint effects of roads and electricity on skilled and unskilled employment and on different occupational sectors. The results are presented in table 3-4 in the Appendix. Access to road and electricity networks increases employment in skilled occupations and decreases employment in unskilled occupations, leading to a structural transformation that consists of a transition to occupations that entail higher skills.

Expanding the connections to the two networks leads to a significant employment increase in skilled and high-skilled occupations, as evidenced by the findings obtained by both data sources, the DHS sample (columns 4 and 5) and the LSMS sample (Column 9). Being 10 km closer to an electricity grid in locations along a main road increases skilled employment by 21.6 percentage points (Column 4) in the DHS sample and high-skilled employment by around 14 percentage points in DHS and 31.2 percentage points in LSMS sample (columns 5 and 9). Similarly, being 10 km closer to a main road in locations along the electricity grid increases skilled employment by 38 percentage points (Column 4) and high-skilled employment by 36 percentage points in the DHS sample (Column 5), and by 21.8 percentage points in the LSMS sample (Column 9).

The joint effect of the two types of infrastructure is positive and highly significant, showing that they play complementary roles in creating skilled jobs. The marginal impact of getting closer to a main road decreases as the distance to the grid increases. Similarly, the marginal impact of getting closer to the grid decreases as the distance to a main road increases. Indeed, being 10 km closer to a main road increases the employment gain that results from being 10 km closer to an electricity grid by 9 percentage points for skilled occupations (Column 4) and by7 percentage points for high-skilled occupations in the DHS sample, and by 8 percentage points for high-skilled occupations in the LSMS sample.

In columns 1, 2, 3, 6, 7, and 8, we show estimates of the effects of roads and electricity on low-skilled employment, including separate effects for farming. Column 1 shows that the probability of self-employment in the farming sector decreases by 22 percentage points in response to being 10 km closer to a main road in locations along the electricity grid (Column 1). The probability of self-employment in the farming sector decreases to an even greater degree, by 43 percentage points in response to being 10 km closer to an electricity grid in areas along a main road (Column 1). Moreover, 10 km closer to a main road and 10 km closer to an electric grid decreases the probability of being self-employed in farming by an additional 13 percentage points.

Columns 2 for the DHS sample and 6 for the LSMS sample consider employment in the farming sector as the dependent variable.[9] In locations along the grid, being 10 km closer to a main road decreases the probability of working as a farmer by 31 percentage points (DHS sample). In locations along a main road, being 10 km closer to the grid decreases the probability of working as a farmer by 8 percentage points (DHS sample). The impact is 9 percentage points when using the LSMS sample. Proximity to roads complements proximity to the electric grid. Farming employment decreases more from being closer to roads for locations closer to the grid, and vice versa. The joint effect of both infrastructures is larger in the DHS than in the LSMS sample. Overall, the findings indicate that these infrastructures reinforce each other in decreasing employment in the agricultural sector.

In columns 3, 7 and 8, we analyze how employment in low-skilled occupations responds to road and electricity expansion. In general, we find large negative effects. Columns 3 and 8 show that the probability of employment in a low-skilled occupation decreases by nearly 5 percentage points in both the DHS and LSMS samples in response to being 10 km closer to a main road in areas along the electric grid. Being 10 km closer to an electric grid in areas along a main road leads the probability of working as a low-skilled employee to decrease to a far greater degree: falling by 21 percentage points in the DHS sample and by 15 percentage points in the LSMS sample. Again, the two types of infrastructure have greater effects in combination. Our analysis shows that being both 10 km closer to a main road and 10 km closer to an electric grid further decreases the probability of being employed in a low-skilled occupation by an additional 6 percentage points in the DHS sample and 2.7 percentage points in the LSMS sample.

These estimates have similar signs as the OLS estimates (table 3-A.4) in most cases. Overall, the findings indicate that improved access to roads and electricity increases the probability of being employed in general (table 3-3) and induces labor reallocation, with workers moving from farming and other low-skilled employment to skilled sectors. Finally, the two types of infrastructure have greater effects in combination.

[9] In column 2, we consider employment in the farming sector as the dependent variable from the DHS. In the LSMS, we also construct a variable that indicates being employed in the farming sector, but this variable does not indicate whether an individual is self-employed or an employee because this information is not provided (Column 6).

3.5.3 Agricultural employment in rural areas

Given agriculture's essential role in rural economies, we now focus on a more detailed analysis of the effects of access to the roads and electricity on agricultural employment. We analyze the separate and combined effects of greater proximity to these networks on agricultural jobs for populations of rural areas, using both the DHS and LSMS samples.

The regression results are presented in table 3-5. They show that road access and electricity significantly decrease agricultural employment in rural areas. In the LSMS sample, being 10 km closer to a main road in a location along the grid decreases the likelihood of working in the farming sector by 1 percentage point (Column 1). Similarly, being 10 km closer to an electric grid in a location along a road decreases agricultural employment by nearly 15 percentage points. The marginal impact of being closer to a main road is smaller when closer to the grid, and vice versa. The estimated effects of these infrastructures are much larger in the DHS sample (Columns 2 and 3).

Improved access to electricity and to roads has a similar impact on the structure of occupations at the local level. Electricity and road access might create new job opportunities, mostly in services, providing a more reliable source of income than working in the agricultural sector. Individuals who have specific skills also have the opportunity to switch from less-skilled employment to more-skilled occupations. Indeed, tables 3-4 and 3-5 implies that people in low-skilled jobs and those working in agricultural occupations (in either the employed or self-employed categories) move to more-skilled occupations in response to greater access to both types of infrastructure. These effects are more pronounced in rural areas.

The OLS estimates reported in Table 3-A.5 confirm all of these results for the IV estimates. Being closer to the roads along with electricity grids decreases employment in the farming sector and vice versa for both DHS and LSMS surveys. The joint OLS effects of these infrastructures are also qualitatively in line with the IV estimates.

The findings suggest that self-employment in agriculture - which represents 25 percent of all jobs in the DHS sample - has declined as infrastructure has expanded. This indicates that Africa is undergoing the same transition - away from agriculture and to other occupations - that has occurred previously in other global regions.

3.6 Heterogeneous effects of roads and electricity on employment

In this section, we analyze the heterogeneity of the effects of greater access to roads and electricity. Our findings show that the impacts vary by level of economic development and by individual characteristics (gender, age, and place of residence). The analysis sheds some light on the possible mechanisms through which investment in infrastructure affects job creation.

We have seen that the effects of roads and electricity vary according to the nature of the occupation. If individual characteristics affect the likelihood of employment in particular sectors, then investments in infrastructure will likely have differential effects depending on the characteristics of different individuals. Moreover, cultural and normative constraints on labor mobility that are prevalent in certain societies, such as those in sub-Saharan Africa, imply that investment in greater access to roads and electricity is likely to have lower impacts among members of certain groups (such as women, for example), depending on how binding such constraints are.

3.6.1 Economic development

We analyze how the effects of roads and electricity on employment differ by level of economic development. Theoretically, heterogeneous impacts could be driven by differences in the occupational choice between countries that differ in the level of economic development. Agricultural occupations, for instance, might be more prevalent in less developed economies, whereas the manufacturing and service sectors might be more present in more developed economies. We classify countries according to their 2019 GDP per capita (table 3-A.2).

The results are reported in columns 7 and 8 of tables 3-6 and 3-7 in the Appendix. The effects of road expansion on employment differ by level of economic development. In areas with high access to electric grids, increasing proximity to the road network by 10 km marginally increases employment probability by 21.2 percentage points in more-developed countries in the LSMS sample and by 37 percentage points in the DHS sample. In less-developed countries, the corresponding effects are 2.7 percentage points (LSMS) and 23.7 percentage points (DHS), respectively. It follows that the effect of greater investments in road networks tends to be larger for more-developed economies. However, the differential effect is smaller in the DHS sample, which, likely offers a more accurate picture because of the greater number of countries included in it (26 countries) than in the LSMS sample (6 countries).

In areas with high access to road networks, being 10 km closer to an electric grid increases the probability of working by 36.1 percentage points in more-developed countries, compared to 7.7 percentage points for less-developed countries in the LSMS sample. In the DHS sample, the corresponding estimates are 4.7 versus 5.4 percentage points, respectively. Again, the effect tends to be larger in more-developed economies.

In locations far enough from road and electricity networks, increasing proximity to both has a positive effect, but this effect is stronger in more-developed countries. These findings imply that the complementary role of roads and electricity is stronger in more-developed African economies.

Results from the OLS regressions (reported in Appendix table 3-A.6), confirm these findings. Corresponding coefficients have similar signs, although the OLS estimates are smaller. Overall, the findings show that investment in infrastructure promotes job creation and that the impacts are larger in more economically advanced African countries. This differential effect tends to validate that the nature of jobs varies across countries with different levels of development and that the expansion of roads and electricity provides fewer benefits to the sectors that predominate in less-developed economies and greater benefits to sectors that predominate in more-developed countries.

3.6.2 Place of residence

We examine how the employment impacts of expanding road and electricity access differ for urban and rural areas. Results are reported in columns 1 and 2 of tables 3-6 and 3-7 in the Appendix. There are net employment gains in response to road expansion in both urban and rural areas in the DHS sample. The marginal employment effect of roads is larger in rural areas. Estimates from the DHS sample show that increasing proximity to the road network by 10 km marginally increases employment by 8.8 percentage points in urban areas, compared to 15.3 percentage points in rural areas along the grid. Similarly, in areas along a main road, expanding electricity access has a larger effect in rural areas; bringing an electric grid 10 km closer to areas with close road networks marginally increases the probability of working by 20 percentage points in rural areas and by only 3 percentage points in urban areas. Additionally, the joint, marginal effect of road and electricity access is positive and larger in rural areas. The rationale for such a result could simply be that rural areas have a larger percentage of unemployed laborers among their residents than is the case in urban areas; thus, the expansion of infrastructure creates employment opportunities to a greater degree for this unemployed and underemployed rural labor force.

Results based on the LSMS sample differ. While the estimated effects of roads and electricity are positive in rural areas (as in the DHS), the marginal effect of roads and the joint effect of roads and electricity are negative. We should note that this discrepancy could be explained by the fact our LSMS sample is drawn mostly from rural areas. The representation of urban areas might not be sufficient. In Appendix table 3-A.6, we present the OLS results, which reinforce the idea that roads and electricity play complementary roles in job creation in rural areas. In contrast to the IV estimates, the OLS estimates show complementarity for infrastructure investment impacts in urban areas.

We also analyze the differential effects of roads and electricity on low-skilled and high-skilled employment in urban and rural areas. The results are reported in table 3-8 in the Appendix. The findings from the LSMS and DHS samples show that low- and high-skilled employment respond differently to the expansion of infrastructure. Low-skilled employment decreases in rural areas but tends to increase in urban areas. By contrast, high-skilled employment tends to increase in urban and rural areas. In areas along the grid, bringing road networks 10 km closer to a rural location marginally decreases the likelihood of working 43 percentage points in the DHS sample (column 4); in urban areas, bringing road networks 10 km closer increases the likelihood of working in a low-skilled occupation by 10 percentage points in the DHS sample (column 3). Similarly, in areas along a main road, bringing an electric grid 10 km closer in rural areas marginally decreases the likelihood of working in a low-skilled occupation by 22.5 percentage points in the DHS sample. The joint effect of roads and electricity is negative in rural and urban areas but is not significant for urban areas.

There are significant employment gains in high-skilled employment in response to the expansion of roads and electricity in urban and rural areas, with few exceptions. In rural areas connected to electricity, bringing roads 10 km closer marginally increases this type of employment by 17.4 percentage points in the LSMS sample and 18.9 percentage points in the DHS sample. In urban areas, the effect of increasing the proximity of roads 10 km is positive in the LSMS sample (9 percentage points and 5 percentage points in the DHS sample. However, the result is not significant for the DHS. In rural areas with high access to roads, bringing electric grids 10 km closer marginally increases the probability of high-skilled employment by over 100 percentage points in the LSMS sample and by 3.6 percentage points in the DHS sample. In urban areas, the corresponding effects are 15.9 (LSMS) and 2.4 (DHS) percentage points, respectively. In urban and rural areas, the joint effect of the expansion of the two types of infrastructure is positive and statistically significant, except in the DHS sample. These findings are consistent with the OLS estimates (Table 3-A.7).

The findings imply that in urban areas, investment in road and electricity infrastructure promotes job creation in both low- and high-skilled occupations. However, in rural areas, there is clear occupational mobility in response to the expansion of roads and electricity, with residents moving from low-skilled to high-skilled jobs. This latter finding is consistent with the negative effect of these two types of infrastructure expansion on agricultural employment (shown in Table 3-5) in rural areas.

3.6.3 Gender

In this section, we examine the different effects on the employment of men and women that stem from road investments and electricity. We run separate regressions for each gender group and present the results in columns 3 and 4 of table 3-6 in the Appendix. We find that the direction in which these infrastructure investments affect employment does not differ by gender. However, the magnitude of the effects does differ.

The DHS sample results show that roads and electricity have greater marginal effects on women than on men. Columns 3 and 4 show that in locations along the grid, bringing roads 10 km closer marginally increases employment by 7.9 percentage points among men and 43 percentage points among women, with these effects being statistically significant. The additional employment gain due to the joint investment in bringing roads and electricity 10 km closer is larger for women (11.6 percentage points) than for men (2.7 percentage points).

Unlike in the DHS sample, the impacts from using the LSMS sample of countries are relatively similar between men and women.

The OLS results are reported in the Appendix Table 3-A.6. The main conclusions from these results qualitatively mirror those from the IV estimates, with a few differences. The OLS results show that electricity expansion increases employment both for men and women and continues to increase employment in areas with high access to roads; this is consistent with the findings from the IV regressions. The OLS results show that, while the joint effect of roads and electricity is positive for both groups, its magnitude is larger for women.

The findings imply that employment gains from investment in road and electricity infrastructure in Africa accrue more among women than among men. However, in certain countries, the effects might be larger for men, as shown in the LSMS. Africa is transitioning from low-paid jobs towards skilled and better-paid occupations; our results show that this transition differs by gender.

3.6.4 Age group

To analyze the differential employment effects of roads and electricity by age, we categorize individuals into age groups: 20-45 and 46-66 years old. Because of differences in occupational choices made by individuals in different age groups, the employment effect of infrastructure expansion might differ by age.

The IV results are reported in columns 5 and 6 of Tables 3-6 and 3-7. For both samples, the results show greater net employment gains for the 20-45 age group than for the 45-66 age group. In locations with high access to electricity, bringing road networks 10 km closer increases the likelihood of working to a greater degree among those in the younger group than those in the older group (+20 vs. +13.5 in the DHS sample, and +5.0 vs. 3.5 in the LSMS). In locations that have high access to roads, increasing proximity to an electric grid by 10 km marginally increases employment by 2.8 percentage points in the younger age group and 36 percentage points in the older age group in the LSMS sample; the corresponding estimates for the DHS sample are 18.6 percentage points for the younger age group and -4.6 percentage points for the older age group. The joint effect of roads and electricity is positive and statistically significant for both age groups in the LSMS sample, with a greater magnitude for the older group. By contrast, in the DHS sample, the effect is significant only among younger people.

The OLS results are reported in Table 3-A.6 and largely mirror the findings from the IV results, with a few differences. However, overall, both the OLS and the IV estimates show that expanded access to roads and electricity affects employment mostly among young and middle-aged individuals. These findings, therefore, suggest that younger individuals can benefit more from expanding these types of infrastructure, most likely because people in these groups can more easily change occupations.

3.7 Robustness checks and evidence in support of identification assumptions

In this section, we present several tests to probe the robustness of our results and provide evidence in support of our identification assumptions. We focus on the main results of the effect of access to roads and electricity on employment using the DHS data set. First, we alternate the choices of clustering of the standard errors. We compute standard errors clustered at the country level, allowing unrestricted auto-correlation among observations in the same administrative unit. We also compute two-way clustered standard errors by country $\times$ year, thus additionally allowing for an arbitrary cross-correlation in the error terms among observations in the same country and year but in different administrative units. Finally, following Conley (1999), we account for spatial correlations within 100 kilometres and a lag of 5 years.[10] Colella et al. (2019) provide a command which accounts for spatial correlation for the instrumental variables. Table 3-9 reports the results for different clustering levels for the standard errors. We note that lower administrative-level clustering increases the standard errors in columns (1) to (4). Accounting for the spatial correlation does not change the standard errors considerably. Overall, the coefficients on proximity to the road and the interaction term between proximity to the road and proximity to the electricity grids remain significant for both the OLS and the IV estimations. The estimated effects of proximity to the electricity grids are only significant for the IV regressions.

We next analyze the validity of our instrumental variable. One concern about proximity to our constructed hypothetical roads as an instrument for proximity to the road network is that these hypothetical roads might be correlated with early (or pre-colonial) development, which might predict employment today. To test this idea, we analyze how proximity to hypothetical roads correlates with key early development indicators, including the intensity of slave trades in an ethnic group, the presence of a city in an ethnic homeland in 1400, and the population density of an ethnic group during the colonial era. Ethnic-level data on slave exports for 17 sub-Saharan African countries come from Nunn and Wantchekon (2011). Nunn (2008) is the first to study the long-term economic impacts of Africa's slave trades. He found that the slave trades that occurred over a period of more than 400 years had a significant negative effect on long-term economic development. [11]. Nunn and Wantchekon (2011) examines the long-term impact of slave trades on trust. While the intensity of slave trades is measured at the country level in Nunn (2008), Nunn and Wantchekon (2011) measures it at the ethnic group level. The latter study also constructs measures of precolonial prosperity that can affect later development and job creation in a location. Using data from Chandler (1987) on the location of African cities with more than 20,000 inhabitants in the year 1400, they construct a variable equal to one if there was a city on the land inhabited by each ethnic group and zero if not.

Finally, We consider population density or urbanization rates during the colonial era.

[10] We use two common STATA commands: $OLS\_Spatial - HAC$ by Hsiang (2010) and reg2hdfespatial by Fetzer (2015).

[11] Between 1400 and 1900, the African continent experienced four simultaneous slave trades. The largest and most well-known is the trans-Atlantic slave trade where, beginning in the 15th century, slaves were shipped from West Africa, West Central Africa and Eastern Africa to the European colonies in the New World. The three other slave trades – the trans-Saharan, Red Sea and Indian Ocean slave trades – are much older and predate the trans-Atlantic slave trade. The beginning of the trans-Saharan and Red Sea slave trades dates back to at least 600 AD (Lovejoy et al. (2003)). During the trans-Saharan slave trade, slaves were taken from the south of the Saharan desert and to Northern Africa. In the Red Sea slave trade, slaves were taken from inland the Red Sea and shipped to the Middle East and India. In the Indian Ocean slave trade, slaves were taken from Eastern Africa and shipped either to the Middle East and India or to plantation islands in the Indian Ocean.

Acemoglu et al. (2002) argue that urbanization affected the strategies undertaken by the colonizers and later development level at specific locations. Nunn and Wantchekon (2011) use population data from Murdock (1959) to measure early twentieth-century population in an ethnic homeland. We used the same data in the analysis.

Table 3-10 presents the correlation between proximity to the ethnic hypothetical roads and these three early development indicators. We control for all the other variables in our main tables. Results do not show any significant correlation between our instrumental variable for the proximity to the main roads and pre-historic prosperity variables. It implies that proximity to the ethnic roads does not reflect the effect of the early development, which is reassuring.

Another concern related to the choice of lightning density and land elevation as instruments for proximity to the electricity grids is that these variables might be correlated with local economic development indicators that predict employment. In Africa, especially in the rural areas, the elevation of land or lightning density may directly affect agricultural outcomes such as growing seasons or gross production that have direct impact on agricultural productivity and agricultural employment growth Dinkelman (2011); Manacorda and Tesei (2020). To test this hypothesis, we collect data on the number of months of the growing season and each location's economic production (evaluated in the dollar). Data on the number of months of the growing season and data on local economic production is provided by DHS geographic data set at the DHS clustering level. In table 3-11, we examine if our instruments are correlated with these indicators of local economic development. Results confirm no correlation between local economic status and our geographic choice of instruments for the proximity to the electricity grids. The findings in tables 3-10 and 3-11 therefore suggest that our instruments satisfy the exclusion restriction and are valid.

3.8 Conclusions

Using two sources of geo-referenced household surveys and spatial data on road and electricity expansion in sub-Saharan Africa, we analyze the average and heterogeneous impacts on job creation that stem from road investments and electrification. We examine the impacts that stem from expanding access to road networks and electric grids, both in isolation and in combination. We find that significant, positive complementarity effects emerge between both types of investments. The positive impacts are heterogeneous across countries, locations, gender, and age groups. The gains from investing in roads and electricity grids tend to be greater in more-advanced economies, where the complementarities are greater. Within countries, the benefits of expanding the reach of such infrastructure differ between urban and rural areas. Employment gains from roads accrue to a greater degree in rural areas. Moreover, the complementarity between the two types of infrastructure is much stronger in rural areas. At the same time, expanding such infrastructure leads to a decline in low-skilled employment, including in agricultural jobs, as people move into other types of work requiring higher skills. In general, though both women and men benefit from better access to roads and electricity, women benefit more. The employment gains from better infrastructure access tend to be greater for younger individuals. Our findings imply that roads and electricity play important roles in the structural transformation of sub-Saharan Africa. The continent is undergoing the same transition from low-productivity agriculture to other occupations as other parts of the world.

Figure 3-1: Change in Sub-Saharan Africa's main roads between 1971-2012

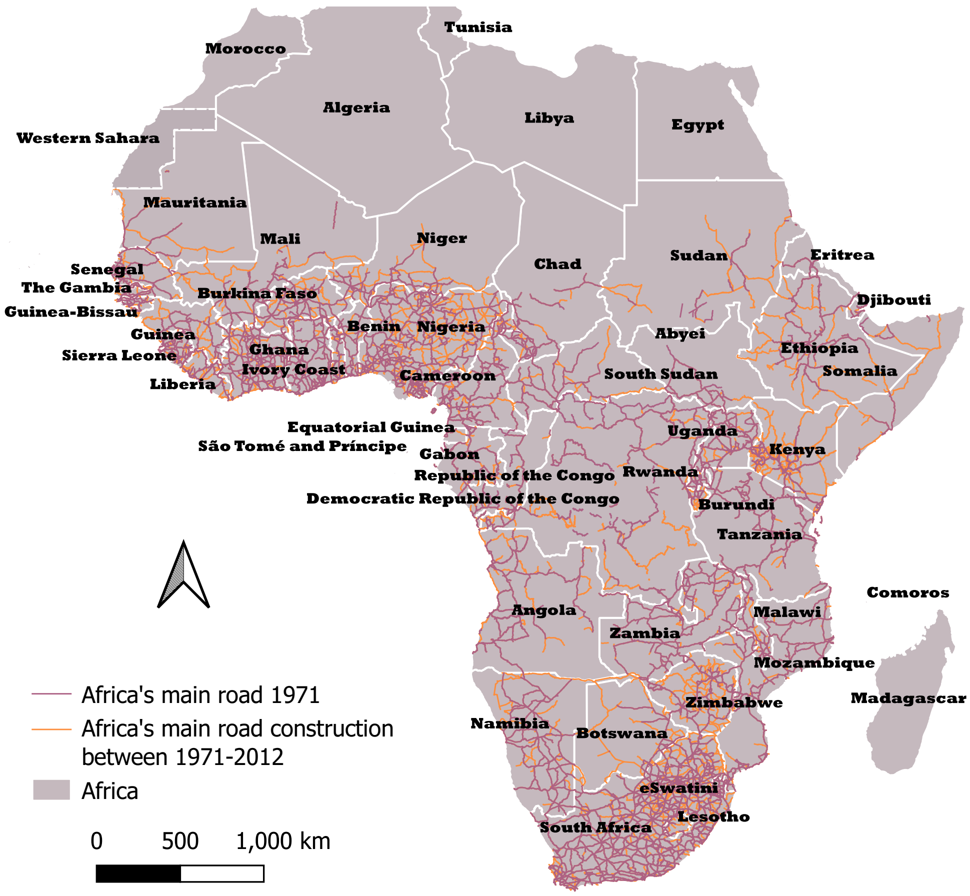


Figure 3-2: Employment pattern and proximity to infrastructure (based on LSMS)

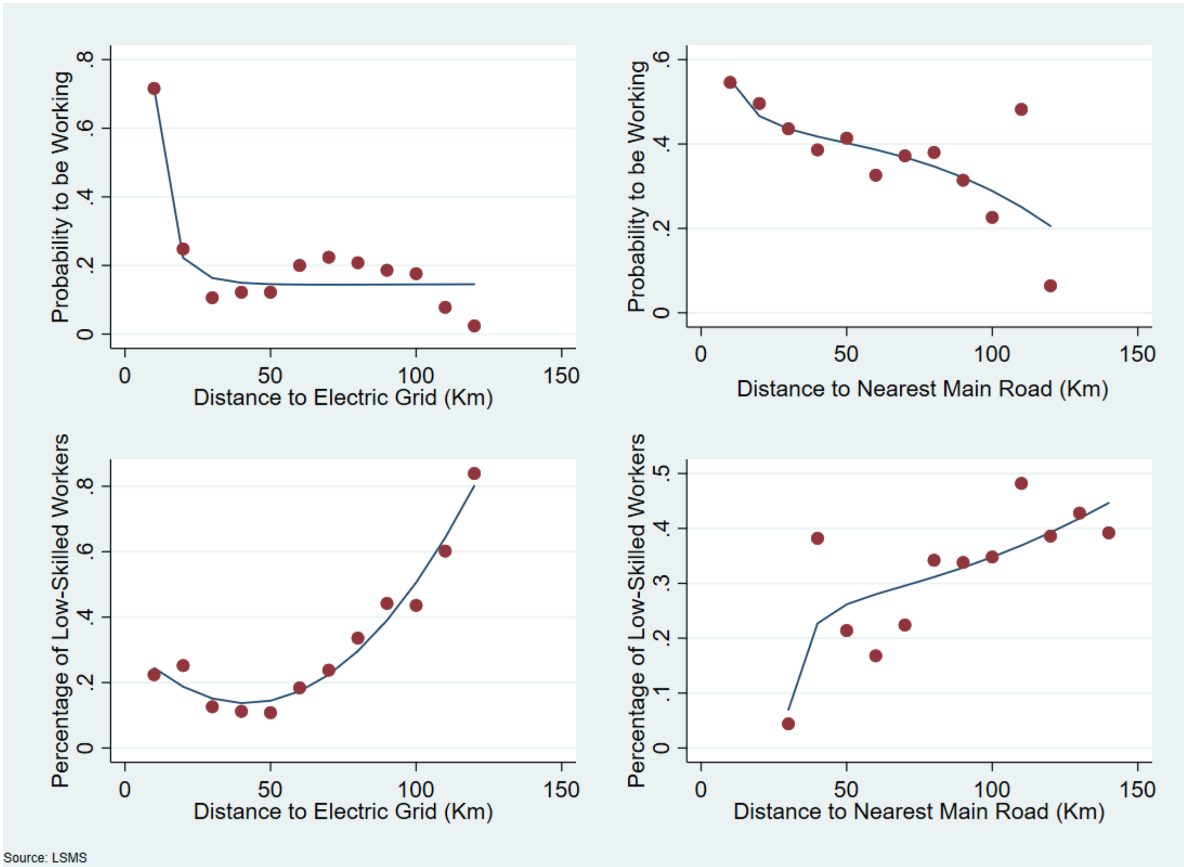


Figure 3-3: Employment pattern and proximity to infrastructure (based on DHS)

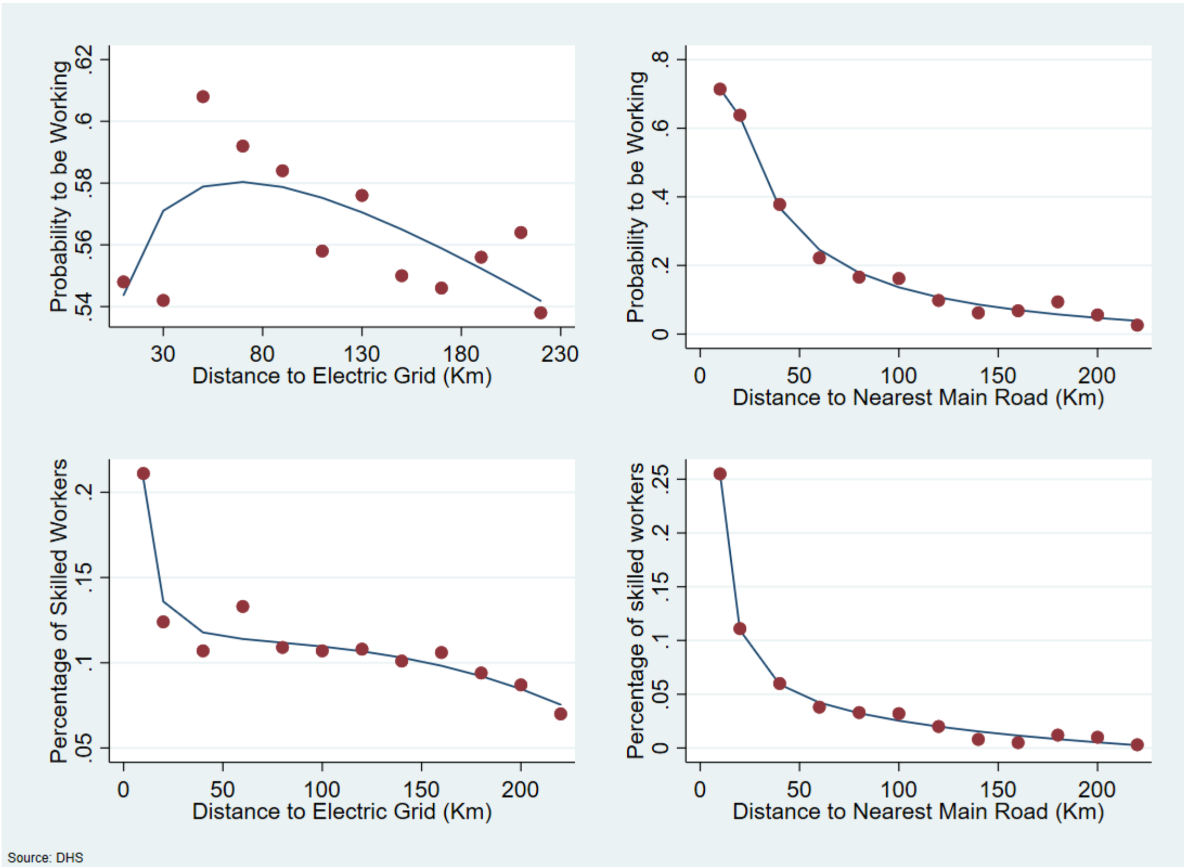


Figure 3-4: Networks of hypothetical roads (green) and of the main roads that existed (as of 2014) (red) in three countries in west Africa

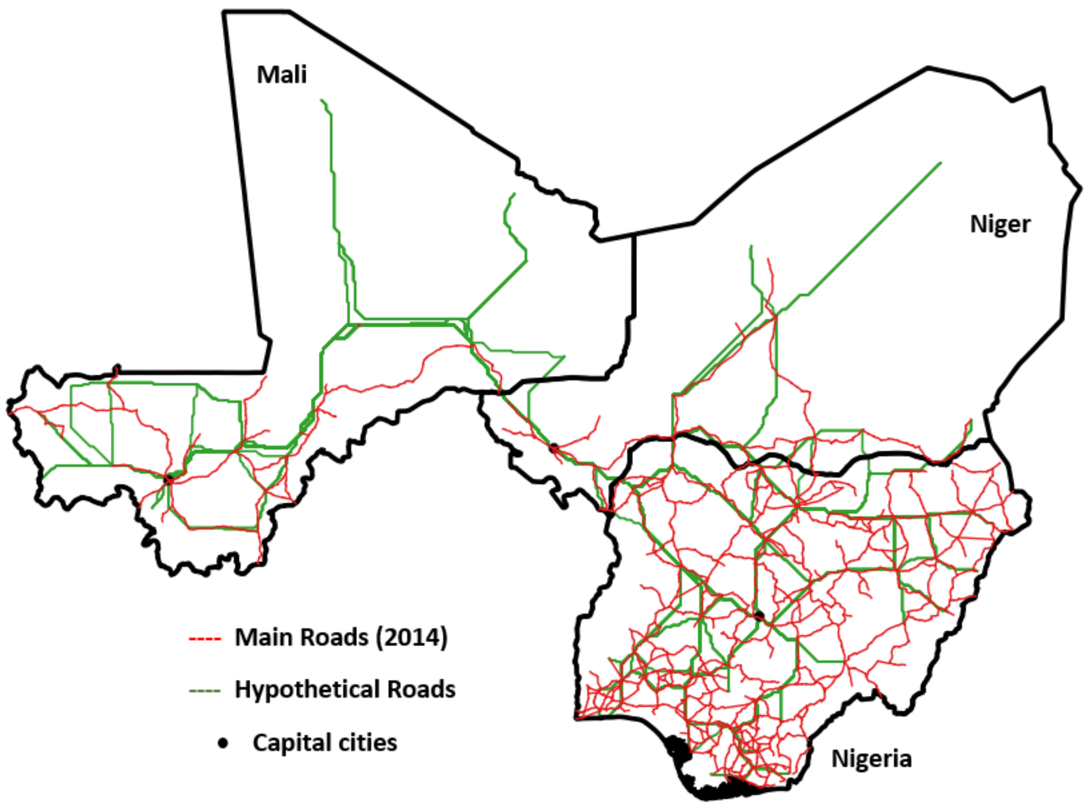


Figure 3-5: The map of lightning density in Africa

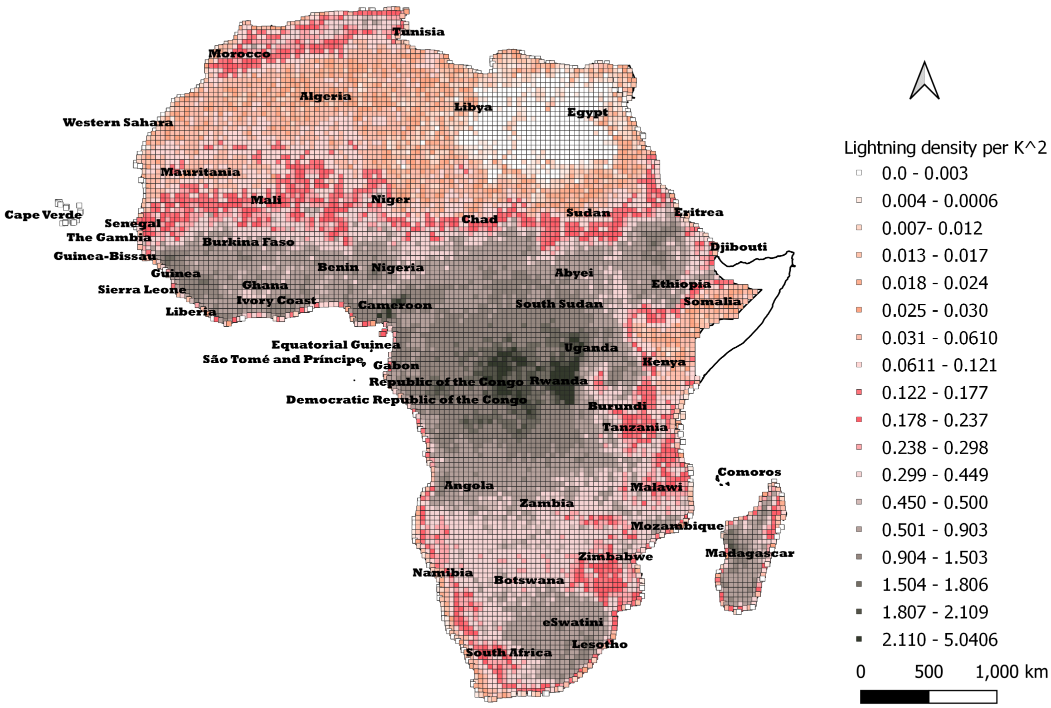


Figure 3-6: Marginal impact of proximity to the roads (left) and electricity (right) on employment

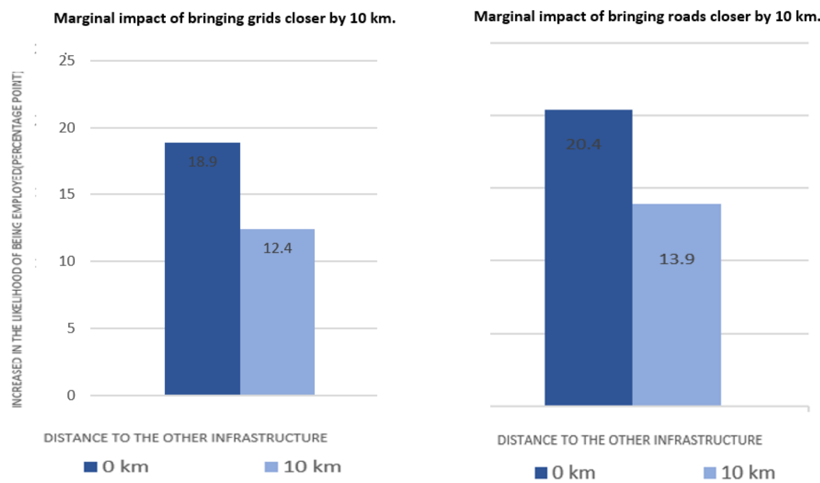


Figure 3-7: Marginal impact of proximity to roads (left) and electricity (right)on hours working during last 7 days

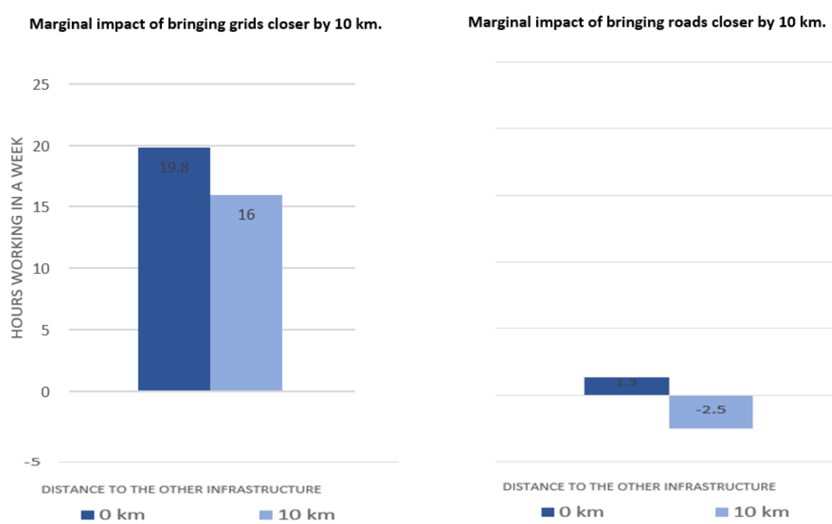


Table 3-1: Descriptive statistics

LSMS					
Variables	Observations	Mean	Std. Dev.	Min	Max
Proximity to road	1,135,766	-15.436	20.339	-192.616	-0.003
Proximity to electric grid	1,135,766	-8.629	19.662	-226.921	-.0000467
Interaction term	1,135,766	266.3476	1154.843	0.0000519	17144.38
Proximity to ethnic road	1,017,060	-38.182	141.244	-14129.83	-0.021
Proximity to the sea	1,135,766	-561.574	298.315	-1393.063	-3.114
Elevation	1,130,271	1037.93	744.5967	0	3451
Lightning	1,135,751	0.455	0.280	0.000355	1.886
Ethnic road * elevation	1,014,139	-37062.46	59870.8	-4889321	0
Currently working	1,135,766	0.588	0.492	0	1
Hours worked	485,213	2.193187	10.32975	0	120
Farmer	891,219	0.5037	0.499	0	1
Low skilled worker	1,135,766	0.494	0.499	0	1
High skilled worker	1,074,871	0.636	0.481	0	1
Age	1,135,766	36.128	12.089	20	65
Female individual	1,135,766	0.519	0.499	0	1
Rural residence	1,135,766	0.861	0.346	0	1
No education	1,135,766	0.230	0.421	0	1
Primary education	1,135,766	0.165	0.371	0	1
Higher education	1,135,766	0.425	0.494	0	1
More developed countries	1,135,766	0.0285	0.166	0	1
DHS					
Variables	Observations	Mean	Std. Dev.	Min	Max
Proximity to road	814,394	-9.978	17.544	-211.506	-0.0000708
Proximity to electric grid	814,394	-26.199	44.327	-332.199	-0.000216
Interaction term	814,394	564.284	2233.378	1.40E-07	46875.13
Proximity to ethnic road	814,394	-32.608	32.500	-221.840	-0.001
Proximity to the sea	804,091	-451.258	353.270	-1738.288	-0.00244
Elevation	798,314	713.656	640.595	-92	3979
Lightning	812,063	0.548	0.378	0	5.024
Ethnic road * elevation	798,314	-22790.94	34388.13	-384234.9	6539.948
Currently working	814,394	0.692	0.461	0	1
self employed farmer	797,570	0.240	0.427	0	1
Farmer employee	797,570	0.0613	0.240	0	1
Farmer	797,570	0.302	0.459	0	1
Low skilled worker	814,394	0.466	0.499	0	1
Skilled worker	814,394	0.133	0.339	0	1
High skilled worker	814,394	0.188	0.391	0	1
Age	814,394	32.345	8.969	20	64
Female individual	814,394	0.754	0.431	0	1
Rural residence	814,394	0.629	0.483	0	1
No education	814,394	0.344	0.475	0	1
Primary education	814,394	0.309	0.462	0	1
Higher education	814,394	0.347	0.476	0	1
More developed countries	814,394	0.440	0.496	0	1

Note: Summary statistics are provided for the largest sample used in the regressions.

Table 3-2: First stage regressions

	Proximity to nearest main road	Proximity to electric grids	Interaction term	Proximity to nearest main road	Proximity to electric grid	Interaction term
	LSMS			DHS		
Proximity to hyp. roads	0.000336***	0.0000352*	-0.0148***	0.0246***	0.237***	-4.635***
	(0.0000326)	(0.0000149)	(0.00138)	(0.00222)	(0.00633)	(0.181)
Elevation	0.0357***	-0.195***	0.376*	-1.320***	-2.854***	61.71***
	(0.00745)	(0.0438)	(0.181)	(0.0558)	(0.183)	(4.063)
Lightning	-0.119***	-0.0648***	-0.458	0.321***	-0.149*	-15.26***
	(0.0144)	(0.0135)	(0.330)	(0.0236)	(0.0614)	(2.250)
Hyp. road*Elevation	-0.000346***	0.0000725	0.00793***	-0.000848***	-0.0556***	1.988***
	(0.0000718)	(0.00000400)	(0.00234)	(0.0000158)	(0.00446)	(0.112)
N	1,014,105	1,014,105	1,014,105	905,246	875,756	875,566
Cragg-Donald Wald F	109.983	109.225	108.882	44.65	44.65	44.65
Windmeijer F	275.44	243.37	740.22	398.98	94.17	142.38
Age	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓
Proximity to the sea	✓	✓	✓	✓	✓	✓
Type of place of residence	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓
Temperature	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

Notes: Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grids. Proximity to the sea is the nearest distance of an individual's geo-location to the sea border. Proximity to the hypothetical road is the nearest distance between a centroid of an individual's cluster and hypothetical roads that connect the exogenous ethnic homeland's location, derived from the map of African ethnolinguistic divisions produced by the American anthropologist George Peter Murdoch in 1959 (Murdock (1959)) to identify the ethnic homelands of each country, to each country's capital city. Proximity to the sea is the nearest distance of an individual's geo-location with the sea border. Lightning density is the yearly mean of the number of lightning strikes, cloud to earth, per square kilometre in the closest location of an individual residence. The interaction term is the interaction between proximity to the road and the elevation of the land. * p<0.1, ** p<0.05, *** p<0.001.

Table 3-3: IV effects of roads and electricity on employment

	Currently working		Hours worked during last 7 days
	LSMS	DHS	LSMS
Proximity to nearest main road	0.00276***	0.0204***	0.133*
	(0.000415)	(0.00294)	(0.0751)
Proximity to electric grids	0.00785***	0.0189***	1.986***
	(0.00137)	(0.00188)	(0.266)
Interaction term	0.000351***	0.000655***	0.0384***
	(0.0000404)	(0.0000623)	(0.00676)
N	1,014,139	785,680	475,640
Cragg-Donald Wald F	64.42	47.11	22.41
Sanderson-Windmeijer F test	191.87	107.04	91.50
P value	0.00	0.00	0.00
Hansen J statistic F	11.567	34.961	143.36
Proximity to the sea	✓	✓	✓
Age	✓	✓	✓
Gender	✓	✓	✓
Type of place of residence	✓	✓	✓
Education	✓	✓	✓
Temperature	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geo-location to the sea border. The coefficient on the "Proximity to nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures it's marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$.

Table 3-4: IV effects of roads and electricity on employment by sector of occupation

	(1) Farming self employed	(2) Farming employee and self employed	(3) low skill worker	(4) Skilled worker	(5) High skill worker	(6) Farming	(7) Low skilled worker	(8) Farming and low skill workers	(9) High skill worker
	DHS					LSMS			
Proximity to road	-0.0221	-0.0310***	-0.0049	0.0380***	0.0361***	-0.00638***	-0.000184	-0.00460***	0.0218***
	(0.0277204)	(0.00397)	(0.0126957)	(0.00757)	(0.00554)	(0.000259)	(0.000364)	(0.000402)	(0.000661)
Proximity to electric grids	-0.0431***	-0.00870***	-0.0207***	0.0216***	0.0142***	0.00915***	-0.00877***	-0.0147***	0.0312***
	(0.0165801)	(0.00234)	(0.0076113)	(0.00453)	(0.00331)	(0.000874)	(0.00125)	(0.00137)	(0.00488)
Interaction term	-0.0013*	-0.000538***	-0.0006*	0.000959***	0.000739***	-0.000153***	-0.000275***	-0.000476***	0.000876***
	(0.0007097)	(0.0000968)	(0.000325)	(0.000187)	(0.000137)	(0.0000261)	(0.0000372)	(0.0000401)	(0.000125)
N	546,241	546,241	546,241	546,241	547,073	868,322	1,014,139	1,014,139	953,244
Cragg-Donald Wald F	60.928	21.3	52.655	60.928	56.932	95.596	68.27	99.541	44.951
Windmeijer F	192.24	34.08	167.88	192.24	176.33	280.80	291,788	220.477	68.732
P value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hansen J statistic F	44.992	21.20	33.807	43.625	33.950	270.404	311.449	333.48	403.088
P value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proximity to the sea	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-5: IV effects of roads and electricity on farming in rural areas

	Farming	Farming self-employed	Farming employee and employed
	LSMS	DHS	
Proximity to road	-0.00129**	-0.0735***	-0.0600***
	(0.000411)	(0.0148)	(0.0109)
Proximity to electric grids	-0.0152***	-0.0275**	-0.0157**
	(0.00250)	(0.00850)	(0.00634)
Interaction term	-0.000409***	-0.00128***	-0.000929***
	(0.0000677)	(0.000306)	(0.000227)
N	894,042	345,110	345,110
Cragg-Donald Wald F	24.04	37.13	36.3
Windmeijer F	86.11	91.79	19.25
P value	0.00	0.00	0.00
Proximity to the sea	✓	✓	✓
Baseline controls	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures the marginal effect of proximity to the electricity grids for locations within zero km of the road.* p<0.1, ** p<0.05, *** p<0.001

Table 3-6: Heterogeneity: IV effects of roads and electricity on employment

	Currently working (DHS)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Urban area	Rural area	male	female	20-45 years	45-66 years	Less developed	More developed
	Currently working (DHS)							
Proximity to road	0.00888	0.0153*	0.00796**	0.0430***	0.0200***	0.0135**	0.0237***	0.0370***
	(0.00655)	(0.00726)	(0.00259)	(0.00829)	(0.00319)	(0.00442)	(0.00297)	(0.00145)
Proximity to electric grids	0.00309***	0.0200***	0.00528***	0.0318***	0.0186***	-0.00469	0.00544***	0.00477***
	(0.000599)	(0.00392)	(0.00118)	(0.00542)	(0.00179)	(0.00355)	(0.000494)	(0.000305)
Interaction term	0.000191***	0.000502***	0.000207***	0.00116***	0.000643***	0.0000347	0.000209***	0.00000710
	(0.0000516)	(0.000142)	(0.0000275)	(0.000197)	(0.0000618)	(0.000111)	(0.0000322)	(0.0000202)
N	293,161	492,519	197,318	588,362	661,504	62,937	525,740	259,940
Cragg-Donald Wald F	32.01	5.505	17.186	12.460	39.120	13.343	62.553	105.326
Windmeijer F	80.34	13.62	74.92	29.58	117.31	45.24	208.45	215.22
P value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proximity to the sea	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-7: Heterogeneity: IV effects of roads and electricity on employment (continue)

	Currently working (LSMS)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Urban area	Rural area	Male	Female	20-45 years	45-66 years	Less developed	More developed
Proximity to road	-0.00289**	0.00297***	0.00229***	0.00305***	0.00483***	0.00351*	0.00271***	0.0212***
	(0.00110)	(0.000437)	(0.000542)	(0.000649)	(0.000503)	(0.00171)	(0.000418)	(0.00605)
Proximity to electric grids	0.00161*	0.0150***	0.00757***	0.00710***	0.00285*	0.0360***	0.00774***	0.0361**
	(0.000701)	(0.00266)	(0.00190)	(0.00182)	(0.00121)	(0.00622)	(0.00136)	(0.0118)
Interaction term	-0.0000365	0.000524***	0.000340***	0.000321***	0.000232***	0.00111***	0.000346***	0.00318**
	(0.0000310)	(0.0000719)	(0.0000542)	(0.0000565)	(0.0000356)	(0.000187)	(0.0000398)	(0.000974)
N	120,097	894,042	487,673	526,466	781,371	232,768	997,779	16,360
Cragg-Donald Wald F	62.46	24.04	35.68	31.26	72.84	32.12	112.32	33.25
Windmeijer F	144.22	86.11	95.73	116.54	216.80	69.08	285.70	66.84
P value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proximity to the sea	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* p<0.1, ** p<0.05, *** p<0.001

Table 3-8: IV effects of roads and electricity on employment by type of occupation and by place of residence

	Low skilled worker			
	Urban area	Rural area	Urban area	Rural area
	LSMS		DHS	
Proximity to road	0.0100***	-0.00508***	-0.00249**	-0.0430***
	(0.00189)	(0.000354)	(0.000689)	(0.0105)
Proximity to electric grids	0.00269**	-0.00245	-0.000877**	-0.0225***
	(0.00104)	(0.00189)	(0.000400)	(0.00602)
Interaction term	0.00000959	-0.000129***	-0.00000570	-0.000876***
	(0.0000397)	(0.0000219)	(0.0000495)	(0.000218)
N	115,528	872,931	194,326	328,900
Cragg-Donald Wald F	61.84	34.25	37.49	37.4
Windmeijer F	89.48	85.66	91.72	101.4
P value	0.00	0.00	0.00	0.00
	High skilled worker			
	Urban area	Rural area	Urban area	Rural area
	LSMS		DHS	
Proximity to road	0.00918***	0.0174***	0.00510	0.0189***
	(0.00162)	(0.00176)	(0.00598)	(0.00396)
Proximity to electric grids	0.0159***	0.143***	0.00241***	0.00368
	(0.00123)	(0.0170)	(0.000613)	(0.00232)
Interaction term	0.000143***	0.00346***	0.0000285	0.000265**
	(0.0000347)	(0.000406)	(0.0000502)	(0.0000842)
N	114,617	812,947	194,326	328,900
Cragg-Donald Wald F	69.19	48.78	44.20	36.11
Windmeijer F	114.25	16.65	97.22	105.26
P value	0.00	0.00	0.00	0.00
Proximity to the sea	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* p<0.1, ** p<0.05, *** p<0.001

Table 3-9: Alternative clustering for DHS data

	(OLS)			
	(1)	(2)	(3)	(4)
	Baseline	Clustered (country level)	Clustered (country-year level)	Accounting for spatial correlation
Proximity to road	0.00103***	0.00103***	0.00103***	0.00103***
	(0.000149)	(0.000220)	(0.000245)	(0.000152)
Proximity to electric grids	0.0000338	0.0000338	0.0000338	0.0000338
	(0.0000529)	(0.0000661)	(0.0000692)	(0.0000551)
Interaction term	0.0000130***	0.0000130***	0.0000130***	0.0000130***
	(0.00000191)	(0.00000222)	(0.00000274)	(0.00000211)
N	814,389	814,389	814,389	814,389
	(IV)			
Proximity to road	0.0204***	0.0204***	0.0204***	0.0204***
	(0.00294)	(0.00333)	(0.00350)	(0.00305)
Proximity to electric grids	0.0189***	0.0189***	0.0189***	0.0189***
	(0.00188)	(0.00225)	(0.00261)	(0.00200)
Interaction term	0.000655***	0.000655***	0.000655***	0.000655***
	(0.0000623)	(0.0000710)	(0.0000755)	(0.0000661)
N	785,680	785,680	785,680	785,680
Proximity to the sea	✓	✓	✓	✓
Age	✓	✓	✓	✓
Gender	✓	✓	✓	✓
Type of place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Temperature	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road. We refer to two common commands: $OLS-Spatial-HAC$ by Hsiang (2010) and reg2hdfespatial by Fetzer (2015). Colella et al. (2019) Provided a command which accounts for spatial correlation for the instrumental variables. * p<0.1, ** p<0.05, *** p<0.001

Table 3-10: Early development variables and proximity to the MerdoK hypothetical roads

	(1)	(2)	(3)
Average slave export	0.000391		
	(0.0101)		
Existence of city among ethnics groups		-0.126	
		(0.123)	
Clonial population density			-1.002
			(1.016)
N	392,458	388,537	338,807
Proximity to the sea	✓	✓	✓
Age	✓	✓	✓
Gender	✓	✓	✓
Type of place of residence	✓	✓	✓
Education	✓	✓	✓
Temperature	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

Notes: The table reports the correlation between early development components and ethnic road diffusion among 17 African countries. The average slave export is calculated based on (Atlantic+Indian)/area km^2. Atlantic and Indian were two major sources of slave export. The existence of a city among ethnic groups is a dummy variable equal to one if that location had a population of more than 20,000 inhabitants in the year 1400. population density is the logarithm of the early twentieth-century population from Murdock (1959). Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

Table 3-11: Effect of flash density and elevation of land on economics variable

	Growing season (month)	Gross production
Flash density ($\hat{2}$km)	0.0100	0.000523
	(0.0278)	(0.00356)
Elevation of land	0.000529	0.0000340
	(0.000364)	(0.0000614)
N	377,009	385,281
Proximity to the sea	✓	✓
Age	✓	✓
Gender	✓	✓
Type of place of residence	✓	✓
Education	✓	✓
Temperature	✓	✓
Country FE	✓	✓
Year FE	✓	✓

Notes: The table reports the correlation between economic growth indicators and ethnic choice of instruments for proximity to the electricity grids among 17 African countries. The growing seasons change between zero to 12 months in each location. Gross production is computed based on thousand dollars in each location of the DHS data set. Robust standard errors in parentheses clustered at the county level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$.

3-A Appendix

Table 3-A.1: Surveys used in the study

		Survey-year	
Num	Country	**DHS**	**LSMS**
1	Angola	2015	
2	Benin	2001, 2012, 2017	
3	Burkina Faso	2003, 2010	
4	Burundi	2010, 2016	
5	Congo	2007, 2013	
6	Ethiopia	2005, 2011, 2016	2011, 2014, 2015
7	Gabon	2012	
8	Ghana	2003, 2008, 2014	
9	Guinea	2005, 2012, 2018	
10	Kenya	2003, 2008, 2014	
11	Liberia	2007, 2013	
12	Lesotho	2004, 2009, 2014	
13	Mali	2001, 2006, 2012, 2018	
14	Malawi	2000, 2004, 2010, 2015	2005, 2010, 2013, 2017, 2019
15	Mozambique	2011, 2015	
16	Niger		2011, 2014
17	Nigeria	2003, 2008, 2013, 2018	2010, 2015, 2018
18	Namibia	2000, 2006, 2013	
19	Sierra Leone	2008, 2013	
20	Senegal	2005, 2010, 2015, 2016	
21	Chad	2014	
22	Togo	2013	
23	Tanzania	2010, 2016	2008, 2010, 2011, 2012, 2014, 2016
24	Uganda	2001, 2006, 2011, 2016	2005, 2009, 2010, 2011, 2013
25	South Africa	2017	
26	Zambia	2016	
27	Zimbabwe	2005, 2010, 2015	

Table 3-A.2: Sub-Saharan Africa: Nine highest-income countries

Countries	GDP per capita (USD) -2019
Gabon	7,767.00
South Africa	6,001.40
Namibia	4,957.50
Angola	2,790.70
Nigeria	2,229.90
Ghana	2,202.10
Kenya	1,816.50
Zambia	1,305.10
Lesotho	1,118.10

Table 3-A.3: OLS: effects of roads and electricity on employment

	Currently working		Hours worked during last 7 days
	LSMS	DHS	LSMS
Proximity to nearest main road	0.0000427*	0.00103***	0.0191***
	(0.0000255)	(0.000149)	(0.00104)
Proximity to electric grids	0.000169***	0.0000338	-0.0211***
	(0.0000318)	(0.0000529)	(0.00419)
Interaction term	0.00000699***	0.0000130***	0.0000735
	(0.000000616)	(0.00000191)	(0.000107)
N	1,135,766	814,389	484,673
Proximity to the sea	✓	✓	✓
Age	✓	✓	✓
Gender	✓	✓	✓
Type of place of residence	✓	✓	✓
Education	✓	✓	✓
Temperature	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-A.4: OLS: effects of roads and electricity on employment by sector of employment

	Farming self employed	Farming employee and self employed	low skill worker	Skilled worker	High skill worker	Farming Farming	Low skilled worker	Farming and low skill workers	High skill worker
	DHS					LSMS			
Proximity to road	-0.00237***	-0.00336***	-0.000297***	0.00136***	0.00130***	-0.000389***	-0.0000362	-0.000126***	0.000168***
	(0.0000628)	(0.0000673)	(0.0000617)	(0.0000504)	(0.0000552)	(0.0000139)	(0.0000187)	(0.0000192)	(0.0000197)
Proximity to electric grids	-0.000417***	-0.000614***	0.0000388*	0.000265***	0.000282***	0.000311***	-0.000485***	0.000132***	0.000230***
	(0.0000164)	(0.0000173)	(0.0000177)	(0.0000147)	(0.0000159)	(0.0000267)	(0.0000223)	(0.0000244)	(0.0000166)
Interaction term	-0.0000115***	-0.0000180***	-0.000000217	0.00000768***	0.00000812***	-0.00000559***	-0.00000259***	-0.00000269***	0.00000347***
	(0.000000522)	(0.000000538)	(0.000000491)	(0.000000390)	(0.000000434)	(0.000000674)	(0.000000351)	(0.000000354)	(0.000000279)
N	556,729	556,729	557,597	557,597	557,597	891,219	1,135,766	1,135,766	1,074,871
Proximity to the sea	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-A.5: OLS effects of roads and electricity on farming in rural areas

	Farming	Farming self-employed	Farming employee and employed
	LSMS	DHS	
Proximity to road	-0.0000480*	-0.00166***	-0.00245***
	(0.0000211)	(0.0000523)	(0.0000742)
Proximity to electric grids	-0.000392***	-0.000255***	-0.000645***
	(0.0000255)	(0.0000180)	(0.0000243)
Interaction term	-0.00000113**	-0.0000115***	-0.0000164***
	(0.000000432)	(0.000000460)	(0.000000608)
N	978,255	545,314	351,700
Proximity to the sea	✓	✓	✓
Baseline controls	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-A.6: OLS: Heterogeneity effects of roads and electricity on employment

	Urban area	Rural area	Male	Female	20-45 age group	45-66 age group	Less developed countries	More developed countries
	Currently working (LSMS)							
Proximity to road	0.000872***	0.000552***	0.000447***	0.000756***	0.000595***	-0.000377	0.000989***	0.00161***
	(0.0000764)	(0.0000294)	(0.0000390)	(0.0000383)	(0.0000309)	(0.000584)	(0.0000194)	(0.0000163)
Proximity to electric grids	0.00271***	0.00328***	0.00322***	0.00334***	0.00328***	0.0000858	0.00221***	0.00213***
	(0.0000768)	(0.0000379)	(0.0000519)	(0.0000445)	(0.0000397)	(0.0000620)	(0.0000210)	(0.0000288)
Interaction term	0.0000346***	0.0000234***	0.0000244***	0.0000293***	0.0000247***	0.00000242	0.0000277***	0.0000211***
	(0.00000139)	(0.000000745)	(0.000000921)	(0.000000929)	(0.000000755)	(0.00000129)	(0.000000405)	(0.000000209)
N	157,511	978,255	546,296	589,470	876,158	259,608	1,103,325	32,441
	Currently working (DHS)							
Proximity to road	-0.000305**	0.0000736	0.000432***	0.000221***	0.000318***	0.000458**	0.000169**	0.000319***
	(0.000141)	(0.0000578)	(0.0000845)	(0.0000639)	(0.0000577)	(0.000164)	(0.0000638)	(0.0000935)
Proximity to electric grids	0.0000180	0.0000166	0.00000926	0.0000399**	-0.0000144	0.0000517	0.000122***	0.000210***
	(0.0000248)	(0.0000191)	(0.0000231)	(0.0000187)	(0.0000165)	(0.0000449)	(0.0000181)	(0.0000274)
Interaction term	-0.000000908	-0.00000239***	0.00000184**	0.00000133**	0.00000157***	0.00000183	0.00000212***	0.00000965***
	(0.00000108)	(0.000000478)	(0.000000688)	(0.000000521)	(0.000000468)	(0.00000134)	(0.000000518)	(0.000000714)
N	299,502	504,589	200,445	603,646	677,100	643,34	544,132	259,959
Proximity to the sea	✓	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* $p<0.1$, ** $p<0.05$, *** $p<0.001$

Table 3-A.7: OLS: effects of roads and electricity on employment by type of occupation and by place of residence

	Low skilled worker			
	Urban area	Rural area	Urban area	Rural area
	LSMS		DHS	
Proximity to road	0.000319***	-0.000505***	-0.000421**	-0.000616***
	(0.0000317)	(0.0000200)	(0.000184)	(0.0000673)
Proximity to electric grids	-0.000520***	-0.0000325	-0.0000147	0.0000331
	(0.0000355)	(0.0000259)	(0.0000317)	(0.0000213)
Interaction term	-0.00000299***	-0.00000798***	0.000000541	-0.00000447***
	(0.000000449)	(0.000000391)	(0.00000139)	(0.000000546)
N	157,511	978,255	198,081	335,221
	High skilled worker			
	Urban area	Rural area	Urban area	Rural area
	LSMS		DHS	
Proximity to road	0.0000672***	0.000182***	0.00183***	0.000974***
	(0.0000160)	(0.0000227)	(0.000180)	(0.0000585)
Proximity to electric grids	0.000329***	0.000242***	0.000191***	0.000302***
	(0.0000436)	(0.0000177)	(0.0000311)	(0.0000179)
Interaction term	0.00000291***	0.00000314***	0.00000833***	0.00000773***
	(0.000000444)	(0.000000318)	(0.00000135)	(0.000000463)
N	156,600	918,271	198,081	335,221
Proximity to the sea	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Proximity to the main road is the nearest distance between an individual's location and the main road in that location. Proximity to the electricity grid is the nearest distance between an individual location and the electricity grid. Proximity to the sea is the nearest distance of an individual's geolocation to the sea border. The coefficient on the "Proximity to the nearest road" measures its marginal effect for locations within zero km of the electric grid. Similarly, the coefficient on the variable "Proximity to electric grids" measures its marginal effect for locations within zero km of the road.* p<0.1, ** p<0.05, *** p<0.001

Bibliography

Abbasi, M., Lebrand, M. S. M., Mongoue, A. B., Pongou, R., and Zhang, F. (2022). Roads, electricity, and jobs: Evidence of infrastructure complementarity in sub-saharan africa. Technical report, World Bank Group, Policy Research Working Paper no. WPS 9976,.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics*, 117(4):1231–1294.

Acemoglu, D. and Robinson, J. A. (2000). Political losers as a barrier to economic development. *American Economic Review*, 90(2):126–130.

Adamopoulou, E. (2013). New facts on infidelity. *Economics Letters*, 121(3):458–462.

Addoum, J. M. (2017). Household portfolio choice and retirement. *Review of Economics and Statistics*, 99(5):870–883.

Agadjanian, V. (2001). Religion, social milieu, and the contraceptive revolution. *Population Studies*, 55(2):135–148.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D. D. (2011). Financial counseling, financial literacy, and household decision-making. *Financial literacy: Implications for retirement security and the financial marketplace*, 181.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life cycle and implications for regulation. *Brookings Papers on Economic Activity*, 2009(2):51–117.

Aggarwal, S., Giera, B., Jeong, D., Robinson, J., and Spearot, A. (2018). Market access, trade costs, and technology adoption: evidence from northern tanzania. Technical report, National Bureau of Economic Research.

Aghion, P. and Durlauf, S. N. (2005). *Handbook of economic growth*. Elsevier.

Agnew, J. and Szykman, L. (2011). Annuities, financial literacy and information overload. *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, pages 158–178.

Agrawal, D. R. (2021). The internet as a tax haven? *American Economic Journal: Economic Policy*, 13(4):1–35.

Agüero, J. M. and Bharadwaj, P. (2014). Do the more educated know more about health? evidence from schooling and hiv knowledge in zimbabwe. *Economic Development and Cultural Change*, 62(3):489–517.

Aklin, M., Bayer, P., Harish, S., and Urpelainen, J. (2017). Does basic energy access generate socioeconomic benefits? a field experiment with off-grid solar power in india. *Science Advances*, 3(5).

Almenberg, J. and Dreber, A. (2015). Gender, stock market participation and financial literacy. *Economics Letters*, 137:140–142.

Alsan, M. M. and Cutler, D. M. (2013). Girls' education and hiv risk: evidence from uganda. *Journal of health economics*, 32(5):863–872.

Alvehag, K. and Soder, L. (2010). A reliability model for distribution systems incorporating seasonal variations in severe weather. *IEEE Transactions on Power Delivery*, 26(2):910–919.

Amorim, G., Lima, R. C., and Sampaio, B. (2018). Broadband internet and protests: Evidence from the occupy movement. *Political Anthropology eJournal*.

Andersen, T. B., Bentzen, J., Dalgaard, C.-J., and Selaya, P. (2011). Does the internet reduce corruption? evidence from us states and across countries. *The World Bank Economic Review*, 25(3):387–417.

Andersen, T. B., Bentzen, J., Dalgaard, C.-J., and Selaya, P. (2012). Lightning, it diffusion, and economic growth across us states. *Review of Economics and Statistics*, 94(4):903–924.

Andersen, T. B. and Dalgaard, C.-J. (2013). Power outages and economic growth in africa. *Energy Economics*, 38:19–23.

Anderson, S. (2018). Legal origins and female hiv. *American Economic Review*, 108(6):1407–39.

Andrianaivo, M. and Kpodar, K. (2012). Mobile phones, financial inclusion, and growth. *Review of Economics and Institutions*, 3(2):30.

Angrist, J. and Imbens, G. (1995). Identification and estimation of local average treatment effects.

Arcand, J.-L. and Wouabe, E. D. (2010). Teacher training and hiv/aids prevention in west africa: Regression discontinuity design evidence from the cameroon. *Health Economics*, 19(S1):36–54.

Arderne, C., Zorn, C., Nicolas, C., and Koks, E. (2020). Predictive mapping of the global power system using open data. *Scientific data*, 7(1):1–12.

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

Azomahou, T. T., Boucekkine, R., and Diene, B. (2016). Hiv/aids and development: A reappraisal of the productivity and factor accumulation effects. *American Economic Review*, 106(5):472–77.

Baker, S., Fradkin, A., et al. (2011). *What drives Job Search?: Evidence from Google Search Data*. Stanford Inst. for Economic Policy Research.

Baral, S., Beyrer, C., Muessig, K., Poteat, T., Wirtz, A. L., Decker, M. R., Sherman, S. G., and Kerrigan, D. (2012). Burden of hiv among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet infectious diseases*, 12(7):538–549.

Barber, B. M. and Odean, T. (2001). The internet and the investor. *Journal of Economic Perspectives*, 15(1):41–54.

Barber, B. M. and Odean, T. (2002). Online investors: do the slow die first? *The Review of financial studies*, 15(2):455–488.

Barman-Adhikari, A. and Rice, E. (2011). Sexual health information seeking online among runaway and homeless youth. *Journal of the Society for Social Work and Research*, 2(2):89–103.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2016). Highways, market access and urban growth in China. SERC Discussion Papers 0200, Spatial Economics Research Centre, LSE.

Bayer, P. J., Bernheim, B. D., and Scholz, J. K. (2009). The effects of financial education in the workplace: Evidence from a survey of employers. *Economic Inquiry*, 47(4):605–624.

Behrens, K., Gaigne, C., Ottaviano, G. I., and Thisse, J.-F. (2006). Is remoteness a locational disadvantage? *Journal of Economic Geography*, 6(3):347–368.

Bellou, A. (2015). The impact of internet diffusion on marriage rates: evidence from the broadband market. *Journal of Population Economics*, 28(2):265–297.

Berg, C., Blankespoor, B., and Selod, H. (2018). Roads and rural development in Sub-Saharan Africa. *Journal of Development Studies*, 54(5):856–874.

Berg, C., Deichmann, U., Liu, Y., and Selod, H. (2017). Transport policies and development. *Journal of Development Studies*, 53(4):465–480.

Berhan, Y. and Berhan, A. (2015). A meta-analysis of risky sexual behaviour among male youth in developing countries. *AIDS research and treatment*, 2015.

Bernheim, B. D., Garrett, D. M., and Maki, D. M. (2001). Education and saving:: The long-term effects of high school financial curriculum mandates. *Journal of public Economics*, 80(3):435–465.

Bertocchi, G. and Dimico, A. (2019). The long-term determinants of female hiv infection in africa: The slave trade, polygyny, and sexual behavior. *Journal of Development Economics*, 140:90–105.

Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324.

Bhuller, M., Havnes, T., Leuven, E., and Mogstad, M. (2013). Broadband internet: An information superhighway to sex crime? *Review of Economic Studies*, 80(4):1237–1266.

Billari, F. C., Giuntella, O., and Stella, L. (2019). Does broadband internet affect fertility? *Population studies*, 73(3):297–316.

Blankespoor, B., Bougna, T., Garduno Rivera, R., and Selod, H. (2017). Roads and the geography of economic activities in mexico. *World Bank Policy Research Working Paper*, (8226).

Blimpo, M. P., Harding, R., and Wantchekon, L. (2013). Public investment in rural infrastructure: Some political economy considerations. *Journal of African Economies*, 22(suppl_2):ii57–ii83.

Blow, L., Browning, M., Ejrnæs, M., et al. (2009). *Marriage and consumption*. CAM, Centre for Applied Microeconometrics.

Bluethgen, R., Gintschel, A., Hackethal, A., and Mueller, A. (2008). Financial advice and individual investors' portfolios. *Available at SSRN 968197*.

Bodie, Z., Merton, R. C., and Samuelson, W. F. (1992). Labor supply flexibility and portfolio choice in a life cycle model. *Journal of economic dynamics and control*, 16(3-4):427–449.

Bogan, V. (2008). Stock market participation and the internet. *Journal of Financial and Quantitative Analysis*, 43(1):191–211.

Bollen, J. and Mao, H. (2011). Twitter mood as a stock market predictor. *Computer*, 44(10):91–94.

Bonfatti, R., Gu, Y., and Poelhekke, S. (2019). Priority roads: the political economy of africa's interior-to-coast roads. Tinbergen Institute Discussion Paper TI 2019-006/VIII, Amsterdam and Rotterdam.

Braitstein, P., Brinkhof, M., Dabis, F., Schechter, M., Boulle, A., Miotti, P., Wood, R., Laurent, C., Sprinz, E., Seyler, C., et al. (2006). Mortality of hiv-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet (London, England)*, 367(9513):817–824.

Brandsaas, E. E. (2021). Household stock market participation and exit: The role of home-ownership.

Briggs, J., Cesarini, D., Lindqvist, E., and Östling, R. (2021). Windfall gains and stock market participation. *Journal of Financial Economics*, 139(1):57–83.

Bruhn, M., Leão, L. d. S., Legovini, A., Marchetti, R., and Zia, B. (2016). The impact of high school financial education: Evidence from a large-scale evaluation in brazil. *American Economic Journal: Applied Economics*, 8(4):256–95.

Brunnermeier, M. K. and Nagel, S. (2008). Do wealth fluctuations generate time-varying risk aversion? micro-evidence on individuals. *American Economic Review*, 98(3):713–36.

Buhaug, H. and Rod, J. K. (2006). Local determinants of african civil wars, 1970–2001. *Political geography*, pages 315–335.

Burgess, R. and Donaldson, D. (2012). Railroads and the demise of famine in colonial india. *Unpublished manuscript*.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padró i Miquel, G. (2015). The value of democracy: evidence from road building in kenya. *American Economic Review*, 105(6):1817–51.

Burke, J. and Hung, A. A. (2021). Trust and financial advice. *Journal of Pension Economics & Finance*, 20(1):9–26.

Burlig, F. and Preonas, L. (2016). Out of the darkness and into the light? development effects of rural electrification. *Energy Institute at Haas WP*, 268:26.

Butler, C. (2000). Hiv and aids, poverty, and causation. *The Lancet*, 356(9239):1445–1446.

Calderon, C., Moral-Benito, E., and Serven, L. (2015). Is infrastructure capital productive? a dynamic heterogeneous approach. *Journal of Applied Econometrics*, 30(2):177–198.

Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Fight or flight? portfolio rebalancing by individual investors. *The Quarterly journal of economics*, 124(1):301–348.

Campbell, J. Y. (2006). Household finance. *The journal of finance*, 61(4):1553–1604.

Campbell, J. Y., Viceira, L. M., Viceira, L. M., et al. (2002). *Strategic asset allocation: portfolio choice for long-term investors*. Clarendon Lectures in Economic.

Canning, D., Mabeu, M., and Pongou, R. (2020a). Colonial origins and fertility: Can the market overcome history. *Unpublished Manuscript*.

Canning, D., Mabeu, M., and Pongou, R. (2020b). Colonial origins and fertility: Can the market overcome history? *Unpublished Manuscript*.

Cao, Y. and Liu, J. (2017). Financial executive orientation, information source, and financial satisfaction of young adults. *Journal of Financial Counseling and Planning*, 28(1):5–19.

Cariolle, J. (2018). Telecommunication submarine-cable deployment and the digital divide in sub-saharan africa. *Available at SSRN 3202941*.

Carlsson, H., Larsson, S., Svensson, L., and Åström, F. (2017). Consumer credit behavior in the digital context: A bibliometric analysis and literature review. *Journal of financial counseling and planning*, 28(1):76–94.

Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2019). The abcs of financial education: Experimental evidence on attitudes, behavior, and cognitive biases. *Management Science*, 65(1):346–369.

Case, A. and Paxson, C. (2013). Hiv risk and adolescent behaviors in africa. *American Economic Review*, 103(3):433–38.

Cecil, D. J., Buechler, D. E., and Blakeslee, R. J. (2014). Gridded lightning climatology from trmm-lis and otd: Dataset description. *Atmospheric Research*, 135:404–414.

Chan, B. T. and Tsai, A. C. (2018). Hiv knowledge trends during an era of rapid antiretroviral therapy scale-up: an analysis of 33 sub-saharan african countries. *Journal of the International AIDS Society*, 21(7):e25169.

Chan, J., Ghose, A., and Seamans, R. (2016). The internet and racial hate crime: Offline spillovers from online access. *MIS Quarterly*, 40(2):381–403.

Chandler, T. (1987). *Four thousand years of urban growth: An historical census*. Mellen.

Chandra, A. and Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4):457–490.

Chang, M. L. (2005). With a little help from my friends (and my financial planner). *Social forces*, 83(4):1469–1497.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1):50–58.

Chinn, M. D. and Fairlie, R. W. (2010). Ict use in the developing world: an analysis of differences in computer and internet penetration. *Review of International Economics*, 18(1):153–167.

Chiplunkar, G. and Goldberg, P. K. (2022). The employment effects of mobile internet in developing countries. Technical report, National Bureau of Economic Research.

Choi, H. and Varian, H. (2012). Predicting the present with google trends. *Economic record*, 88:2–9.

Ciacci, R. (2021). A matter of size: Comparing iv and ols estimates.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, 18(2):491–533.

Cole, S., Paulson, A., and Shastry, G. K. (2014). Smart money? the effect of education on financial outcomes. *The Review of Financial Studies*, 27(7):2022–2051.

Colella, F., Lalive, R., Sakalli, S. O., and Thoenig, M. (2019). Inference with arbitrary clustering.

Collier, P. and Sterck, O. (2018). The moral and fiscal implications of antiretroviral therapies for hiv in africa. *Oxford Economic Papers*, 70(2):353–374.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45.

Conley, T. G. and Taber, C. R. (2011). Inference with "difference in differences" with a small number of policy changes. *The Review of Economics and Statistics*, 93(1):113–125.

Cornish, F. and Campbell, C. (2009). The social conditions for successful peer education: a comparison of two hiv prevention programs run by sex workers in india and south africa. *American journal of community psychology*, 44(1-2):123–135.

Cosar, A. K. and Fajgelbaum, P. D. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics*, 8(1):24–56.

Crocker, T. and Williams, T. (2022). Exploring the online banking practices of subsidized housing residents. *Journal of Poverty*, 26(2):141–161.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–74.

Cummings, B. F. and James III, R. N. (2014). Factors associated with getting and dropping financial advisors among older adults: Evidence from longitudinal data. *Journal of Financial counseling and Planning*, 25(2):129.

Curcuru, S., Heaton, J., Lucas, D., and Moore, D. (2004). Heterogeneity and portfolio choice: Theory and evidence, handbook of financial econometrics. *Forth% coming*.

Dallabetta, G. A., Miotti, P. G., Chiphangwi, J. D., Saah, A. J., Liomba, G., Odaka, N., Sungani, F., and Hoover, D. R. (1993). High socioeconomic status is a risk factor for human immunodeficiency virus type 1 (hiv-1) infection but not for sexually transmitted diseases in women in malawi: Implications for hiv-1 control. *Journal of Infectious Diseases*, 167(1):36–42.

Dave Donaldson, R. H. (2016). Railroads and American economic growth: A & quot; Market Access & quot; approach. *The Quarterly Journal of Economics*, 131(2):799–858.

De Neve, J.-W., Fink, G., Subramanian, S., Moyo, S., and Bor, J. (2015). Length of secondary schooling and risk of hiv infection in botswana: evidence from a natural experiment. *The Lancet Global Health*, 3(8):e470–e477.

De Stefano, T., Kneller, R., Timmis, J., et al. (2014). The (fuzzy) digital divide: The effect of broadband internet use on uk firm performance.

De Walque, D. (2007). How does the impact of an hiv/aids information campaign vary with educational attainment? evidence from rural uganda. *Journal of Development Economics*, 84(2):686–714.

De Walque, D. (2009). Does education affect hiv status? evidence from five african countries. *The World Bank Economic Review*, 23(2):209–233.

Deng, F. M. (1997). Ethnicity: an african predicament. *The Brookings Review*, 15(3):28–31.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from south Africa. *American Economic Review*, 101(7):3078–3108.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934.

Dorado, D. A. D., Obille, K. L. B., Garcia, R. P. P., and Olgado, B. S. (2019). Sexual information behaviuor of filipino university students. In *International Conference on Information*, pages 283–289. Springer.

Drexler, A., Fischer, G., and Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2):1–31.

Duflo, E., Dupas, P., and Kremer, M. (2015a). Education, hiv, and early fertility: Experimental evidence from kenya. *American Economic Review*, 105(9):2757–97.

Duflo, E., Dupas, P., and Kremer, M. (2015b). School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from kenyan primary schools. *Journal of public Economics*, 123:92–110.

Dunkle, K. L., Jewkes, R. K., Murdock, D. W., Sikweyiya, Y., and Morrell, R. (2013). Prevalence of consensual male–male sex and sexual violence, and associations with hiv in south africa: A population-based cross-sectional study. *PLoS medicine*, 10(6).

Dupas, P. (2011). Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 3(1):1–34.

Durante, R. (2009). Risk, cooperation and the economic origins of social trust: An empirical investigation. MPRA Paper 25887, University Library of Munich, Germany.

Dyson, T. (2003). Hiv/aids and urbanization. *Population and Development Review*, 29(3):427–442.

Edward Miguel, S. S. and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725–753.

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social media and protest participation: Evidence from russia. *Econometrica*, 88(4):1479–1514.

Epstein, H. and Morris, M. (2011). Concurrent partnerships and hiv: an inconvenient truth. *Journal of the international AIDS society*, 14:1–11.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Fagereng, A., Gottlieb, C., and Guiso, L. (2017). Asset market participation and portfolio choice over the life-cycle. *The Journal of Finance*, 72(2):705–750.

Fagereng, A., Mogstad, M., and Rønning, M. (2021). Why do wealthy parents have wealthy children? *Journal of Political Economy*, 129(3):703–756.

Fajgelbaum, P. D. and Redding, S. J. (2014). External integration, structural transformation and economic development: Evidence from argentina 1870-1914.

Fajgelbaum, P. D. and Schaal, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, 88(4):1411–1452.

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting behavior and the internet. *American Economic Review*, 104(7):2238–65.

Farhi, E. and Panageas, S. (2007). Saving and investing for early retirement: A theoretical analysis. *Journal of Financial Economics*, 83(1):87–121.

FCC (2019). 2019 broadband deployment report. *GN-Docket*, pages 18–238.

Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *The American Political Science Review*, 97(1):75–90.

Federal-reserve (2022). How many americans own stock?

Fenske, J. (2015). African polygamy: Past and present. *Journal of Development Economics*, 117:58–73.

Fenton, L. (2004). Preventing hiv/aids through poverty reduction: the only sustainable solution? *The Lancet*, 364(9440):1186–1187.

Fernandes, A. M., Mattoo, A., Nguyen, H., and Schiffbauer, M. (2019). The internet and chinese exports in the pre-ali baba era. *Journal of Development Economics*, 138:57–76.

Fernandes, D., Lynch Jr, J. G., and Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, 60(8):1861–1883.

Fernández-López, S., Rey-Ares, L., and Vivel-Búa, M. (2018). The role of internet in stock market participation: just a matter of habit? *Information Technology & People*.

Fetzer, T. (2015). Conley spatial hac standard errors for models with fixed effects. *Thiemo Fetzer. http://www. trfetzer. com/conley-spatial-hac-errors-with-fixed-effects (accessed October 17, 2019).*

Fisher, G., Wallace, L., Ryan, L., McCammon, R., and McArdle, J. (2016). Internet usage among older adults: Findings from the health and retirement study 2002-2014. In *GERONTOLOGIST*, volume 56, pages 45–45. OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA.

Fisher, M., Goddu, M. K., and Keil, F. C. (2015). Searching for explanations: How the internet inflates estimates of internal knowledge. *Journal of experimental psychology: General*, 144(3):674.

Fortson, J. G. (2008). The gradient in sub-saharan africa: socioeconomic status and hiv/aids. *Demography*, 45(2):303–322.

Fried, S. and Lagakos, D. (2020). Electricity and firm productivity: A general-equilibrium approach. NBER Working Papers 27081, National Bureau of Economic Research, Inc.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5):1854–1899.

Gall, J., Sabin, K., Frescura, L., Sabin, M. L., Erkkola, T., and Toskin, I. (2017). Global trends of monitoring and data collection on the hiv response among key populations since the 2001 un declaration of commitment on hiv/aids. *AIDS and Behavior*, 21(1):34–43.

Garofalo, R., Herrick, A., Mustanski, B. S., and Donenberg, G. R. (2007). Tip of the iceberg: Young men who have sex with men, the internet, and hiv risk. *American journal of public health*, 97(6):1113–1117.

Georgarakos, D., Inderst, R., and Deuflhard, F. (2015). Financial literacy and savings account returns. Technical report, European Central Bank.

Georgarakos, D. and Pasini, G. (2011). Trust, sociability, and stock market participation. *Review of Finance*, 15(4):693–725.

Gerrans, P. and Hershey, D. A. (2017). Financial adviser anxiety, financial literacy, and financial advice seeking. *Journal of Consumer Affairs*, 51(1):54–90.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to success: The Impact of the golden quadrilateral project for the location and Performance of Indian manufacturing. *Economic Journal*, 126(591):317–357.

Glaser, M. and Klos, A. (2013). Causal evidence on regular internet use and stock market participation. *Available at SSRN 2021158*.

Gonzales, P. and Ajami, N. (2017). Social and structural patterns of drought-related water conservation and rebound. *Water Resources Research*, 53(12):10619–10634.

Grable, J. E. and Joo, S.-h. (2001). A further examination of financial help-seeking behavior. *Journal of Financial Counseling and Planning*, 12(1):55.

Grainger, C. A. and Zhang, F. (2019). Electricity shortages and manufacturing productivity in pakistan. *Energy Policy*, 132:1000–1008.

Gray, P. B. (2004). Hiv and islam: is hiv prevalence lower among muslims? *Social science & medicine*, 58(9):1751–1756.

Greenwood, J., Kircher, P., Santos, C., and Tertilt, M. (2017). The role of marriage in fighting hiv: A quantitative illustration for malawi. *American economic review*, 107(5):158–162.

Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). Iq and stock market trading. *Journal of Finance*, 66(6):2121–2164.

Guiso, L., Haliassos, M., and Jappelli, T. (2003). Household stockholding in europe: where do we stand and where do we go? *Economic Policy*, 18(36):123–170.

Guiso, L. and Jappelli, T. (2005). Awareness and stock market participation. *Review of Finance*, 9(4):537–567.

Guiso, L., Sapienza, P., and Zingales, L. (2004). The role of social capital in financial development. *American economic review*, 94(3):526–556.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic perspectives*, 20(2):23–48.

Guiso, L., Sapienza, P., and Zingales, L. (2008). Trusting the stock market. *the Journal of Finance*, 63(6):2557–2600.

Guldi, M. and Herbst, C. M. (2017). Offline effects of online connecting: the impact of broadband diffusion on teen fertility decisions. *Journal of Population Economics*, 30(1):69–91.

Guner, N. and Knowles, J. (2004). Marital instability and the distribution of wealth. *Unpublished Manuscript. Universidad Carlos III, Department of Economics. Madrid.*

Hallman, K. (2005). Gendered socioeconomic conditions and hiv risk behaviours among young people in south africa. *African Journal of AIDS Research*, 4(1):37–50.

Han, E. and Paik, C. (2017). Ethnic integration and development in china. *World Development*, 93:31–42.

Hanna, S. D. (2011). The demand for financial planning services. *Journal of Personal Finance*, 10(1):36–62.

Herbst, J. (2000). Economic incentives, natural resources and conflict in Africa. *Journal of African Economies*, 9(3):270–294.

Herbst, J. (2014). *States and power in Africa*. Princeton University Press.

Herrera Dappe, M. and Lebrand, M. (2021). Infrastructure and structural change in the horn of africa. Policy Research Working Paper Series WPS 9870, The World Bank.

Hilgert, M. A., Hogarth, J. M., and Beverly, S. G. (2003). Household financial management: The connection between knowledge and behavior. *Fed. Res. Bull.*, 89:309.

Hjort, J. and Poulsen, J. (2019). The arrival of fast internet and employment in africa. *American Economic Review*, 109(3):1032–79.

Hogarth, J. M. and Anguelov, C. (2004). Are families who use e-banking better financial managers? *Journal of Financial Counseling and Planning*, 15(2):61–77.

Hong, H., Kubik, J. D., and Stein, J. C. (2004). Social interaction and stock-market participation. *The journal of finance*, 59(1):137–163.

Hsiang, S. M. (2010). Temperatures and cyclones strongly associated with economic production in the caribbean and central america. *Proceedings of the National Academy of sciences*, 107(35):15367–15372.

Hvide, H. K., Meling, T., Mogstad, M., and Vestad, O. L. (2020). Broadband internet and the stock market investments of individual investors. *Available at SSRN 3751721*.

Igbo, H. I., Egbe-Okpenge, E. G., and Awopetu, R. G. (2013). Influence of information and communication technology on behavior problems of nigerian youths. *Procedia-Social and Behavioral Sciences*, 84:97–106.

Iorio, D. and Santaeulàlia-Llopis, R. (2016). Education, hiv status and risky sexual behavior: how much does the stage of the hiv epidemic matter? *Available at SSRN 2820991*.

Jedwab, R. and Storeygard, A. (2021). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960-2010. *Journal of the European Economic Association*.

Johnson, K. and Way, A. (2006). Risk factors for hiv infection in a national adult population: evidence from the 2003 kenya demographic and health survey. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 42(5):627–636.

Kaiser, T. and Menkhoff, L. (2017). Does financial education impact financial literacy and financial behavior, and if so, when? *The World Bank Economic Review*, 31(3):611–630.

Kam, J., Stowers, K., and Kim, S. (2019). Monitoring of drought awareness from google trends: a case study of the 2011–17 california drought. *Weather, Climate, and Society*, 11(2):419–429.

Karagiannis, G. M., Chondrogiannis, S., Krausmann, E., and Turksezer, Z. I. (2017). Power grid recovery after natural hazard impact. *Joint Research Center: European Union*.

Kebede, H. A. (2020). The gains from market integration: The welfare effects of new rural roads in Ethiopia . Mimeo, University of Virginia.

Kimemia, K. A. and Mugambi, M. M. (2016). Social media and teenage pregnancy among students in secondary schools in imenti north sub-county, meru, county kenya. *International Journal of Scientific Research and Management*, 4(09):4586–4606.

Koletić, G., Štulhofer, A., Tomić, I., and Knežević Ćuća, J. (2019). Associations between croatian adolescents' use of sexually explicit material and risky sexual behavior: A latent growth curve modeling approach. *International Journal of Sexual Health*, 31(1):77–91.

Kramer, M. M. (2016). Financial literacy, confidence and financial advice seeking. *Journal of Economic Behavior & Organization*, 131:198–217.

Ky, S., Rugemintwari, C., and Sauviat, A. (2018). Does mobile money affect saving behaviour? evidence from a developing country. *Journal of African Economies*, 27(3):285–320.

Lall, S. V. and Lebrand, M. (2020). Who wins, who loses? Understanding the spatially differentiated effects of the belt and road initiative. *Journal of Development Economics*, 146(C).

Lebrand, M. (2022). Infrastructure and structural change in the lake chad region. Policy Research Working Paper Series WPS 9899, The World Bank.

Lebrand, M. S. M., Mongoue, A. B., Pongou, R., and Zhang, F. (2023). Does africa need more roads in the digital age? Technical report, World Bank Group, Policy Research Working Paper.

Leung, C. J. (2013). *Determinants of stock market participation among elderly US households with internet access*. PhD thesis.

Liang, P. and Guo, S. (2015). Social interaction, internet access and stock market participation—an empirical study in china. *Journal of Comparative Economics*, 43(4):883–901.

Lovejoy, P. E., Trotman, D. V., and Lovejoy, P. E. (2003). *Ethnic Designations of the Slave trade and the Reconstruction of the History of Trans-Atlantic Slavery*. ProQuest Information and Learning.

Luhrmann, M., Serra-Garcia, M., and Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, 10(3):309–32.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2017). Optimal financial knowledge and wealth inequality. *Journal of Political Economy*, 125(2):431–477.

Lusardi, A. and Mitchell, O. S. (2007). Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of monetary Economics*, 54(1):205–224.

Lusardi, A. and Mitchell, O. S. (2008). Planning and financial literacy: How do women fare? *American economic review*, 98(2):413–17.

Lusardi, A. and Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of economic literature*, 52(1):5–44.

Macdonald, B. and Hülür, G. (2021). Internet adoption in older adults: findings from the health and retirement study. *Cyberpsychology, Behavior, and Social Networking*, 24(2):101–107.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

Magnusson, B. M., Nield, J. A., and Lapane, K. L. (2015). Age at first intercourse and subsequent sexual partnering among adult women in the united states, a cross-sectional study. *BMC public health*, 15(1):98.

Magruder, J. R. (2011). Marital shopping and epidemic aids. *Demography*, 48(4):1401–1428.

Malgouyres, C., Mayer, T., and Mazet-Sonilhac, C. (2019). Technology-induced trade shocks? evidence from broadband expansion in france.

Manacorda, M. and Tesei, A. (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica*, 88(2):533–567.

Martincus, C. V., Carballo, J., and Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. *Journal of Development Economics*, 125:21–39.

McCoy, S. I., Kangwende, R. A., and Padian, N. S. (2010). Behavior change interventions to prevent hiv infection among women living in low and middle income countries: a systematic review. *AIDS and Behavior*, 14(3):469–482.

Meekers, D. and Calves, A.-E. (1997). 'main'girlfriends, girlfriends, marriage, and money: the social context of hiv risk behaviour in sub-saharan africa. *Health Transition Review*, pages 361–375.

Miguel, E. (2005). Poverty and witch killing. *Review of Economic Studies*, 72(4):1153–1172.

Miller, G. (2010). Contraception as development? new evidence from family planning in colombia. *The Economic Journal*, 120(545):709–736.

Minnaar, U., Gaunt, C., and Nicolls, F. (2012). Characterisation of power system events on south african transmission power lines. *Electric power systems research*, 88:25–32.

Moneke, N. (2020). Can big push infrastructure unlock development? evidence from ethiopia. *STEG Theme*, 3:14–15.

Mongoue, B. A. and Pongou, R. (2023). 'son bias and hiv'.

Moran, J. P. (2009). *Teaching sex: The shaping of adolescence in the 20th century*. Harvard University Press.

Morten, M. and Oliveira, J. (2018). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper*, 22158:1–64.

Mu, R. and van de Walle, D. (2011). Rural roads and local market development in vietnam. *Journal of Development Studies*, 47(5):709–734.

Muller-Crepon, C., Hunziker, P., and Cederman, L.-E. (2020). Roads to rule, roads to rebel: Relational state capacity and conflict in africa. *Journal of Conflict Resolution*, 65:002200272096367.

Murdock, G. P. (1959). Africa its peoples and their culture history.

Nelson, A. and Deichmann, U. (2004). African population database documentation. *Retrived from http://na. unep. net/siouxfalls/globalpop/africa/Africa_index. html*.

Nunn, N. (2008). The long-term effects of africa's slave trades. *The Quarterly Journal of Economics*, 123(1):139–176.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in africa. *American Economic Review*, 101(7):3221–52.

Okello Candiya Bongomin, G., Ntayi, J. M., Munene, J. C., and Malinga, C. A. (2018). Mobile money and financial inclusion in sub-saharan africa: the moderating role of social networks. *Journal of African Business*, 19(3):361–384.

Okoye, D., Pongou, R., and Yokossi, T. (2019). New technology, better economy? the heterogeneous impact of colonial railroads in nigeria. *Journal of Development Economics*, 140:320–354.

Ostrom, E. (1990). *Governing the Commons: The evolution of institutions for collective action*. Political Economy of Institutions and Decisions. Cambridge University Press.

Parente, S. L. and Prescott, E. C. (1999). Monopoly rights: A barrier to riches. *American Economic Review*, 89(5):1216–1233.

Paxson, C. H. (1992). Using weather variability to estimate the response of savings to transitory income in Thailand. *American Economic Review*, 82(1):15–33.

Pongou, R. (2009). Anonymity and infidelity: Ethnic identity, strategic cross-ethnic sexual network formation, and hiv/aids in africa. *Working paper, Department of Economics, Brown University.*

Pongou, R. and Serrano, R. (2013). Fidelity networks and long-run trends in hiv/aids gender gaps. *American Economic Review*, 103(3):298–302.

Quinn, T. C., Mann, J. M., Curran, J. W., and Piot, P. (1986). Aids in africa: an epidemiologic paradigm. *Science*, 234(4779):955–963.

Raleigh, C. and Hegre, H. (2009). Population size, concentration, and civil war. a geographically disaggregated analysis. *Political Geography*, 28(4):224–238.

Ratchford, B. T., Talukdar, D., and Lee, M.-S. (2001). A model of consumer choice of the internet as an information source. *International Journal of Electronic Commerce*, 5(3):7–21.

Richardson, B. (2022). How digital technology changed the face of mortgage industry.

Romer, C. and Bernstein, J. (2009). The job impact of the american recovery and reinvestment plan.

Rosenfeld, M. J. and Thomas, R. J. (2012). Searching for a mate: The rise of the internet as a social intermediary. *American Sociological Review*, 77(4):523–547.

Saad, L. (2019). What percentage of americans own guns. *GALLUP. Retrieved September*, 28:2020.

Sabri, M. F. and Aw, E. C.-X. (2019). Financial literacy and related outcomes: The role of financial information sources. *International journal of business and society*, 20(1):286–298.

Samad, H. and Zhang, F. (2018). Electrification and household welfare: evidence from pakistan.

Samad, H. A. and Zhang, F. (2016). Benefits of electrification and the role of reliability: evidence from india. *World Bank policy research working paper*, (7889).

Samad, H. A. and Zhang, F. (2017). Heterogeneous effects of rural electrification: Evidence from bangladesh. *World Bank Policy Research Working Paper*, (8102).

Sanderson, E. and Windmeijer, F. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221.

Schwartz, C. R. (2013). Trends and variation in assortative mating: Causes and consequences. *Annual Review of Sociology*, 39:451–470.

Selod, H. and Soumahoro, S. (2019). Highway politics in a divided government: Evidence from mexico. *World Bank Policy Research Working Paper*, (8710).

Shannon, K., Strathdee, S. A., Goldenberg, S. M., Duff, P., Mwangi, P., Rusakova, M., Reza-Paul, S., Lau, J., Deering, K., Pickles, M. R., et al. (2015). Global epidemiology of hiv among female sex workers: influence of structural determinants. *The Lancet*, 385(9962):55–71.

Skimmyhorn, W. (2016). Assessing financial education: Evidence from boot camp. *American Economic Journal: Economic Policy*, 8(2):322–43.

Statista (2022). Share of population using digital banking in the united states from 2018 to 2022.

Stevenson, B. and Wolfers, J. (2007). Marriage and divorce: Changes and their driving forces. *Journal of Economic perspectives*, 21(2):27–52.

Stolper, O. (2018). It takes two to tango: Households' response to financial advice and the role of financial literacy. *Journal of Banking & Finance*, 92:295–310.

Swanepoel, S. (1998). *Real Estate Confronts Reality*. Dearborn Real Estate Education.

Taaffe, J., Morrill, R. L., and Gould, P. R. (1963). Transport expansion in underdeveloped countries: A comparative analysis. *Geographical Review*, 53:503–52910.

Takyi, B. K. (2003). Religion and women's health in ghana: Insights into hiv/aids preventive and protective behavior. *Social science & medicine*, 56(6):1221–1234.

Tollefsen, A. and Buhaug, H. (2015). Insurgency and inaccessibility. *International Studies Review*, 17:6–25.

Tombe, T. and Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of China. *American Economic Review*, 109(5):1843–1872.

Turkle, S. (2016). *Reclaiming conversation: The power of talk in a digital age*. Penguin.

UNAIDS (2020). Global aids update 2020—seizing the moment: Tackling entrenched inequalities to end epidemics. *Geneva, Switzerland*.

Uzuke, C. and Oyeka, I. (2017). Modelling fertility among polygamous households. *JPSS*, page 247.

Van de Walle, D., Ravallion, M., Mendiratta, V., and Koolwal, G. (2017). Long-term gains from electrification in rural india. *The World Bank Economic Review*, 31(2):385–411.

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial economics*, 101(2):449–472.

Vestman, R. (2019). Limited stock market participation among renters and homeowners. *The Review of Financial Studies*, 32(4):1494–1535.

Vissing-Jørgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of political Economy*, 110(4):825–853.

Wakefield, M. A., Loken, B., and Hornik, R. C. (2010). Use of mass media campaigns to change health behaviour. *The Lancet*, 376(9748):1261–1271.

Wisniowski, B. and Lindstrom, J. (2012). How the world wide web impacts the financial advisor. *Technology Tools for Today's High-Margin Practice: How Client-Centered Financial Advisors Can Cut Paperwork, Overhead, and Wasted Hours*, pages 61–71.

Wohlman, I. M. (2020). The worldwide digital divide and access to healthcare technology. In *Technology and Global Public Health*, pages 241–256. Springer.

Wolla, S. (2017). Evaluating the effectiveness of an online module for increasing financial literacy. *Social Studies Research and Practice*.

World Economic Forum, World Bank, and African Development Bank (2017). The africa competitiveness report 2017 : Addressing africa's demographic dividend.

Wright, P. J. (2011). Mass media effects on youth sexual behavior assessing the claim for causality. *Annals of the International Communication Association*, 35(1):343–385.

Wright, Paul J, S. C. B. A., Johnson, Jennifer A, E., and B, M. (2019). Condom use, pornography consumption, and perceptions of pornography as sexual information in a sample of adult us males. *Journal of health communication*, 24(9):693–699.

Wusu, O. (2013). Exposure to media content and sexual health behaviour among adolescents in lagos metropolis, nigeria. *African journal of reproductive health*, 17(2):157–168.

Xiao, J. J. and Porto, N. (2016). Which financial advice topics are positively associated with financial satisfaction. *Journal of Financial Planning*, 29(7):52–60.

Yang, Y., Fu, W. J., and Land, K. C. (2004). A methodological comparison of age-period-cohort models: the intrinsic estimator and conventional generalized linear models. *Sociological methodology*, 34(1):75–110.

Yao, Y. (2022). Fertility and hiv risk in africa. *Review of Economic Dynamics*, 45:109–133.

Zagorsky, J. L. (2005). Marriage and divorce's impact on wealth. *Journal of Sociology*, 41(4):406–424.

Zhang, F. (2018). *In the dark: how much do power sector distortions cost south Asia?* The World Bank.