Essays on Health and Development Economics

by

Bluette Arcady Mongoue

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

© Bluette Arcady Mongoue, Ottawa, Canada, 2022

Abstract

Chapter 1 – We offer the first evidence of the causal effect of son bias and offspring sex composition on HIV spread. Using individual level data from African countries, we find that a mother is more likely to be infected with HIV if she has a higher natural propensity to bear daughters rather than sons, or if she has fewer sons than desired. This finding is robust across different measures of son bias and sibling sex composition. The effect differs by type of kinship, marriage mode, and level of economic development. We uncover four underlying mechanisms. First, son bias leads to larger family size, thereby increasing the likelihood of unprotected sex. Second, son bias causes wives to be unfaithful. Third, son bias causes husbands to take additional wives when the first wife is more likely to produce daughters. Fourth, women who bear a daughter when unmarried are less likely to find a husband, leading to a larger number of sexual partners throughout their lifetime. In contrast, son preferences have no effect on a male's HIV status or sexual infidelity, except when he has low education. The findings support the classical theory of sex determination according to which child sex is determined by the male sexual gametes, as well as modern generalizations of this theory that also assign a role to mothers.

Chapter 2 – We examine how the advent of fast Internet has affected credit access in sub-Saharan Africa, and document implications for commercial banks' excess liquidity. Exploiting the gradual arrival of submarine Internet cables on the coast of African countries in combination with individual-level data, we apply a difference-in-differences framework and find large positive effects. These effects are larger in urban areas, driven by increased entrepreneurship, employment, productivity, and client solvency. In rural areas, the effects are significantly positive for non-agricultural business owners, and negative for agricultural business owners, which mirrors labor migration out of the agricultural sector to the trading and services sectors. Country-level panel data available for 32 countries in sub-Saharan Africa confirm the main findings, and show that banks' willingness to grant loans has increased and their excess liquidity has significantly decreased in response to the introduction of fast Internet.

Chapter 3 – Are roads and Internet complements or substitutes in the process of job creation? We use individual and firm level datasets from sub-Saharan African countries and an identification technique that combines difference-in-differences and instrumental variables to address this question. We find that roads and Internet jointly positively affect employment, and so act as complements in Sub-Saharan Africa's structural transformation. Skilled employment gains more from the simultaneous expansion of these infrastructures, and their effect also differ by gender, age, educational level, and level of economic development. We also uncover substantial gains for firms.

Declaration of Authorship

I am gratified for Roland Pongou's contribution to the research associated with the three chapters of this thesis. Although this thesis is written by myself, his contribution is just as significant as mine. I am also grateful for the contribution of Mathilde Lebrand and Fan Zhang to the research associated with the third chapter.

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: December 23, 2022

General Introduction

Key determinants of economic development include health, infrastructure, human capital, and equity, among other factors. Given the importance of these different factors, researchers and policymakers have given a lot of attention to the question of what are the appropriate combination of instruments and strategies to foster economic development in less developed countries. I intend to explore this wide-ranging topic in my thesis. My research focuses on three types of incontrovertible factors influencing economic development: son bias and its effect on health, credit access in the digital era, and labour market response to infrastructure accessibility. I study these topics in three different chapters.

In the first chapter, we investigate how preference for sons affects the likelihood of parents to be HIV infected, with a main focus on mothers' infection risk. Son bias is prevalent in sub-Saharan Africa, with a good majority of societies being patrilineal. In most African societies, women who have given birth to a son are more socially valued. However, owing to the fact that biological factors attribute to men the characteristic of carrying the genes responsible for determining the sex of the fetus, we explain why son bias has a greater impact on the health of mothers. This study is the first to establish a causal relationship between son preferences and the serological status of parents. We use individual-level datasets from sub-Saharan African countries to establish this causal relation. We exploit differences in son preferences between individuals within a society and across cultures to demonstrate that son bias increases mothers' likelihood of being infected with HIV. Interestingly, we find that the nature of marital regime affects how son bias affects the likelihood of HIV infection, with a greater impact on women in a monogamous regime. Analyzing the possible drivers of the effect of son bias on HIV infection, we show that fertility decisions, women's infidelity, marriage market outcomes, and sexual behaviour respond to son bias in ways that significantly increase the likelihood of mothers of daughters to be at greater infection risk. We demonstrate that these effects operate because preference for sons promotes discrimination against women. In fact, in societies where son bias is predominant, social pressure gives women who have not given birth to sons an incentive to adopt deviant sexual behavior in the aim of having sons.

In the second chapter, we explore how ICT expansion alleviates bank reluctance to finance sub-Saharan African economies. Specifically, we emphasis on the role of fast Internet on private credit access. Exploiting the gradual arrival of submarine cables along the coast of sub-Saharan African countries, we demonstrate that the differences in credit access between sub-Saharan African countries can be partly attributed to different in Internet access. Our primary goal in this chapter is to examine whether individuals' and firms' accessibility to formal credit is determined by internet accessibility. The rationale is that internet access reduces market imperfections and information asymmetry, which in turn decrease liabilities and the ability of individuals in need of financing to provide collateral (to guarantee loans). These effects operate through an increase in labor force participation and the rise of entrepreneurship initiatives. At the individual, the firm, and the national levels, we estimate statistically significant negative effects of the accessibility to internet on the banks' reluctance to grant loans, with larger effects in urban areas. Thus, economic policies aiming to improve economic development in sub-Saharan Africa should promote access to formal financing through investments that reduce frictions in the financing market.

In the last chapter, we analyze the complementary role of roads and Internet in job creation and structural transformation in sub-Saharan Africa. There is a wealth of evidence that infrastructure investments improve employment outcomes. In spite of this, very little research has been conducted on the interaction between different types of infrastructure and their joint effects on the labour market. This chapter answers this question by investigating the combined effects of road access and internet availability on the supply of, and the demand for, labour force in Sub-Saharan African countries, and how social and demographic factors impact these effects. Using a strategy that combines a difference-in-differences technique and an instrumental variables approach, we compare the employment outcomes of individuals and firms that have a better access to both infrastructures to those with access to only one or neither of the two infrastructures. When roads networks and Internet access are increased in isolation and in combination, findings indicate that both forms of investment have considerable and positive effects, and that their interaction bolster both the demand for, and the supply of, labour. The benefit of investing simultaneously in these infrastructures tends to be greater in more advanced African economies. Within countries, the benefits of expanding these infrastructures vary by location. Roads contribute more to job creation in rural areas than in urban areas, and the Internet has a greater impact in urban area. Additionally, high-speed Internet's greater influence in metropolitan areas enhances the effect of roads. At the same time, as both infrastructures expand, low-skilled employment decreases, particularly agricultural jobs, as individuals migrate to higher-skilled jobs. This chapter suggests that, a major factor in Sub-Saharan Africa's structural development could be the joint development of different types of infrastructures. Therefore, it could be advantageous to integrate infrastructural development into poverty reduction policies that target improvements in the labor market in this region.

Acknowledgments

My dissertation was made possible thanks to God's wisdom, strength, and direction, which resulted in motivation, determination, health, and a successful journey.

My Ph.D. journey would not have been successful without the assistance of several people.

Prior to anything else, I would like to express my deepest gratitude to my Ph.D. supervisor, Professor Roland Pongou. My gratitude cannot be adequately expressed in simple words for the incessant support you have provided and your priceless supervision during this Ph.D. journey. Dynamic, visionary, motivated, and immensely knowledgeable, you have impacted me. In addition, I appreciated your compassion, and support during tough times. Having you as advisor was the best decision I ever made.

I would also like to express my sincere gratitude to the members of the thesis committee. I am especially grateful for the time, encouragement, comments and suggestions of Professors Myra Yazbeck, Louis Philippe Beland and Louis-Philippe Morin.

I also thank the examiners –Professors Susan Godlonton and David Gray– for the time allocated to the evaluation of this thesis, also for their comments and suggestions that will certainly allow to improve this work.

In appreciation of the support and assistance of the faculty members and the administrative staff in the Department of Economics at the University of Ottawa, I would like to thank them as well.

In appreciation of the inspiring discussions on research, their friendship, and the support I received from my fellow doctoral students, I wish to express my sincere gratitude to Mansoureh Abassi, Landry Kuate, Marie-Christelle Mabeu, Philippe Kabore and Jean-Baptiste Tondji.

I would also like to thank my many friends across Canada, the USA, and Cameroon for their encouragement during my Ph.D. journey.

I will certainly not forget to thank my parents and my family, and more particularly my husband and my children who showed great patience with me and who above all gave me support and love during this journey which was not that short and challenging.

Dedication

Dedicated to: My parents: **Justin Potago** and **Martine Djonou**
my husband **Alfred Mongoue Teto** and my **children**.
Thanks to you all, I was able to accomplish all of this.

Contents

List of Tables

List of Figures

Chapter 1

Son Bias and HIV

1.1 Introduction

Son bias–a mindset or cultural mores according to which male children are more valuable than their female counterparts–is a widespread phenomenon that causes parents to desire more sons than daughters. The economics literature shows that son bias affects several important outcomes: fertility choices (Jayachandran & Kuziemko (2011), Leung (1991), Choi & Hwang (2015), Filmer et al. (2008), Larsen et al. (1998), Das Gupta & Bhat (1997)); wife bargaining power in the household (Mughal & Javed (2018), Schneebaum & Mader (2013), L. Li & Wu (2011), Takaku (2018)); relationship and family stability (Dahl & Moretti (2004)); and social and demographic phenomena such as the missing women problem (Sen (1990), R. Pongou (2008), Anderson & Ray (2010), Anderson & Ray (2018), Milazzo (2018), Jayachandran (2017a), Bhalotra & Cochrane (2010), Hu & Schlosser (2015)).

In this paper, we study the causal impact of son bias on HIV outcome in sub-Saharan Africa, which is a question that has not received any attention in the literature. In general, the consequences of son bias in this region have not been widely documented(R. Pongou (2013a), R. Pongou (2020), K. D. B. Pongou R. & Tsala Dimbuene (2017)). This lack of attention could be explained by the fact that son bias does not have the same manifestations in this part of the world as in some countries of South and East Asia, where it is more prevalent.[1] Yet, our data from 37 sub-Saharan African countries show that the ideal gender composition favors one gender for a significant number of individuals, and among those individuals, 80 percent of men versus 60 percent of women desire more sons than daughters (Figure 1-4). However, significant heterogeneity exists, which could be partially explained by the diversity of family systems in African societies (Brodeur et al. (2020), with patrilineal and patrilocal families exhibiting a greater demand for sons.

Several theoretical mechanisms support the thesis that preference for sons could increase the likelihood of getting infected with HIV. One reason is unprotected sex. There is broad evidence for the fact that couples that have only daughters and who display a son bias have a strong incentive to keep producing children until they bear a son. This implies that such couples will engage in unprotected sex and will therefore increase their likelihood of HIV infection, especially if one member of the couple has other sexual partners. We explore several other interesting mechanisms: infidelity within married couples; polygamy; time

[1] For example, while son bias has been shown to result in the abortion of female fetuses and the neglect and high morbidity and mortality of girls in countries such as India and China (Sen (1990), R. Pongou (2008), Anderson & Ray (2010) and Jayachandran & Kuziemko (2011)), no such manifestations have been observed in sub-Saharan African countries (R. Pongou et al. (2017), Mabeu & Pongou (2019)).

spent on the marriage market; and the lifetime number of sexual partners. Importantly, we analyze men and women separately, given that the consequences of not having a son may differ by gender in the African context.

Africa is indeed an ideal setting for our analysis. In addition to displaying significant heterogeneity in son bias, it has the highest prevalence of HIV/AIDS, and this pandemic continues to generate losses in terms of human capital and financial resources. More than 25.6 million people were living with HIV in sub-Saharan Africa in 2018 (UNAIDS (n.d.)), and statistics also reveal there were over 470,000 HIV-related deaths in this region that same year, compared with 770,000 worldwide. We show that son preferences play a significant role in HIV spread in this region.

Information from the World Bank reveals a decrease in the prevalence of HIV in sub-Saharan Africa over the period 2004-2020 (part one of figure (1-2)). Our sample displays a similar trend in the second part of the figure, showing a downward trend in the prevalence of HIV, which is strongly correlated with a decrease in son bias as defined by Jayachandran (2017b). Despite this decrease in the prevalence of HIV in sub-Saharan Africa, we can observe in the same figure, a relatively large and persistent gap between its prevalence in this region and that of other regions of the world. This shows that the question of reducing AIDS in sub-Saharan Africa remains a relevant issue. Interestingly, the decline in the preference for sons (observed in figure (1-2) between 2004 and 2020) is correlated with that of the HIV prevalence in sub-Saharan Africa, which leads us to wonder if economic policies aiming at reducing the prevalence of HIV in this region should integrate into their strategies ways to reduce discrimination against women. This conclusion can be reached only if we show a causal relationship between son bias and HIV infection risk.

The descriptive relationship between Son bias and HIV prevalence is illustrated in Figure (1-5). Our regression-based analysis, based on different measures of siblings sex composition and son bias, confirms that son preferences significantly increase the likelihood of female HIV, but not male HIV.

We analyze a pooled sample of Demographic and Health Surveys from 37 sub-Saharan African countries. These surveys are representative at the national and sub-national level, and most variables are comparable across years and across countries. For our analysis, we compute four different measures of sibling sex composition and son bias: (a) a binary indicator for whether the first child is female; (b) a binary indicator for whether all the children born to a parent are girls; (c) a mother's natural propensity to bear daughters; (d) and a binary indicator for whether the actual proportion of sons is smaller than the ideally desired proportion[2].

We find that each of these variables has a positive and statistically significant effect on the probability of HIV infection among women. However, none increases HIV infection among men[3]. It follows that although fathers are more biased than mothers, son bias primarily increases the likelihood of HIV infection among mothers.

Further combining our data with the ethnographic atlas map produced by Murdock (Murdock (1959)), we use a specification that accounts for ethnic homelands within each country to assess several heterogeneous effects of son bias. We find that son bias is larger in patrilineal societies, where the increase in the odds to bear daughters increases the likelihood of being HIV-infected whereas the effect in matrilineal societies is either negative or statistically nil. In tribes where polygamy is practiced, mothers who had a first child girl are 0.4 percentage points more likely to be HIV infected, in comparison to their counterparts whose first child is a son. Meanwhile, in predominantly monogamous tribes, the effect is lower and statistically nil (0.21 percentage points). We find that while son bias leads fathers of daughters toward polygamy, it leads mothers of daughters toward infidelity. In the case of polygamy, the man is the vector of potential infections, while in the second case, it is the woman who is the vector of possible infections. We would have expected to have a higher effect among monogamous tribes in the event of the mother's infidelity due to the greater vulnerability of women to sexual infections. The greater effect among polygamous tribes could mirror the fact that fathers of daughters are more prone to engage in polygamy than mothers of daughters engaging in extra marital relationships to bear a son.

In societies where bride price is mandatory, the impact of bearing only daughters on the probability of HIV infection is 0.64 percentage points higher. By contrast, in tribes where bride price is a token or optional, the effect of bearing only daughters on HIV is close to zero and statistically insignificant. This allows us to conclude that there is a social cost associated with not bearing sons, which is higher in tribes where bride price is mandatory, due to the patrilineal nature of such tribes. That social cost favors debauched sexual behaviour–in order to have a son–which in turn increases the effect of bearing daughters on HIV infection risk among societies where the bride price is required.

These puzzling findings are further explained through the different mechanisms we ex-

[2] We note that the first two variables are commonly used in studies exploring the effects of son bias; however, the last two variables are new. In particular, a mother's natural propensity to have more daughters than sons is computed as the residual of the regression of a binary indicator for whether one has more daughters than sons on the total number of children. This variable therefore accounts for fertility preferences.

[3] These effects are causal, especially given that in the context of sub-Saharan Africa, sex-selective abortion is non-existent, and therefore the realization of child sex is random given parents' characteristics. Also, all the studies that document the effects of son bias in societies where sex-selective abortion is absent interpret these effects as being causal for this reason.

plore. As a prelude to the analysis of the different mechanisms underlying the effect of sibling sex composition on HIV spread, we document the existence of important heterogeneity in this relationship by level of son bias, age at first birth, and by level of education. Investigating heterogeneity by level of son bias is useful to establish that son bias is really the main driver of this relationship. In our sample, son bias varies across individuals, as some individuals desire more sons than daughters, and others desire more daughters than sons.[4] We find that a greater probability to bear daughters significantly increases the likelihood of getting infected with HIV only in son-biased mothers. Displaying a son bias itself has no effect in mothers who have more sons than daughters. These findings establish that son bias has a real effect on HIV spread.

We also analyze heterogeneity by age at first birth (15-24, 25-39 and 40 years old and up). We find that the effect of son bias on HIV is highest among women who had their first child between the age of 25 and 39. This might be explained by the fact that the demand for children is highest in this age group. Mothers who started having children late (after 40) have a shorter fertility clock and therefore do not have the biological ability to multiply children until they bear a son. Also, mothers who started having children early (between 15 and 24) have a longer fertility clock and may be more optimistic about having a son even if their first child is a girl; they may therefore have a lower incentive to engage in infidelity. For these women, having only daughters gives 0.5 percent more chance to be HIV infected vs 2 percent for the second group of mothers. For men, son bias does not appear to increase HIV in any of the age groups. Studying heterogeneity by level of education, we find that the effect of sibling sex composition and son bias on HIV spread is greater in more educated mothers, but not visible at all more educated fathers. We will argue later that these findings are consistent with the fact that the XX-XY theory of sex determination is only taught in high school in most countries.

We identify four direct channels through which son bias affects an individual's likelihood of being infected with HIV. First, we find that mothers whose first child is a female or who have a greater propensity to bear daughters have significantly more children than other mothers. Moreover, even among mothers who have exceeded their ideal number of children, we find that there is a demand for additional children that is higher among couples who tend to bear more daughters. Consequently, there must be a strong demand for unprotected sex among couples who tend to have more female children, which in turn might explain their higher risk of HIV infection.

Second, we find that married women who have a greater propensity to bear daughters or

[4]Importantly, we note that these two groups of parents have similar offspring sex ratio, showing that being biased toward a particular sex does not actually increases one's probability to bear a child of that sex.

who have fewer sons than desired are more likely to have extramarital relationships. We do not find any such effect among men. These results are consistent with the traditional theory of sex determination–the XX-XY sex-determination system (Stevens (1905), Sutton (1903) Brush (1978), Henking (1891)). According to this theory, the sex of a fetus is primarily determined by the father. During conception, the mother contributes an X chromosome and the father contributes either an X chromosome or a Y chromosome, resulting in a daughter (XX) or a son (XY). This theory implies that a woman has a greater incentive to seek extramarital relationships in order to have a boy when her husband is only giving her daughters or when he is not giving her the number of boys she desires.

Interestingly, given that the XX-XY theory of sex determination is mostly taught in high school in most countries, this can shed light on why the effect of son bias on HIV spread is greater in more educated mothers, but absent in more educated fathers. Indeed, to the extent that knowledge of this theory shapes incentives, a more educated woman has a greater incentive to seek extramarital relationships in order to have a boy when her husband is not giving her the desired number of boys; but an educated father who is unable to bear sons has no such incentive, ”knowing” from theory that engaging in extramarital relationships might not produce the desired outcome. It follows that son bias should affect HIV spread more significantly among educated mothers and less significantly among educated fathers, which is the findings we obtain.

Third, we find that mothers who are more likely to bear daughters than sons are more likely to have a polygamous husband. This finding implies that such mothers are either more likely to find husbands who already have other wives, or that their husbands are more likely to take a second wife if they are the first wife. Indeed, we find that a man whose first wife is more likely to bear daughters than sons is more likely to take another wife, provided that the first wife does not have only daughters.[5] We remark that the traditional XX-XY sex determination theory cannot fully explain why a man is more likely to take on a second wife when his first wife is more likely to bear girls. However, this finding is consistent with the modern theory of sex determination. According to this latter theory, the sex of a child can also be determined by factors affecting the mother (W. H. James (1996)). If men believe this theory, this might lead to polygamy when the first wife is only bearing daughters.

Furthermore, we note that the fact that men choose polygamy over infidelity in response to their first wife not giving them the desired offspring sex ratio reflects the fact that in most

[5]When the first wife has only daughters, that decreases the demand for a second wife by the husband. This could be explained by the fact that the husband is not optimistic about producing a son even with a different wife, as he is only able to give daughters to his first wife. However, if the first wife has at least one son but has a greater propensity to produce girls, then the husband has a greater incentive to take on an additional wife.

African societies, a son born out-of-wedlock cannot become his father's heir or cannot inherit his father's properties. Remark that the effect of son bias on infidelity is only present among women, despite men displaying a greater son bias (Figure 1-4). This might be suggesting that son bias causes a woman to be unfaithful in order to bear a son that will be attributed to her husband, which could help her save her marriage or increase her bargaining power in the household.[6]

The fourth channel we explore focuses on marriage timing for mothers who had their first child when they were single. We find that women who bear a daughter when unmarried are more likely to remain unmarried compared to their counterparts who gave birth to a son.[7] Similarly, we find that married women who had a daughter before marriage spend a longer time on the marriage market compared to their counterparts who had a son. Consistent with these findings, we show that having a daughter when unmarried leads to a higher lifetime number of sexual partners, which increases the likelihood of HIV infection.

The remaining of this paper is organized as follows. Section 2 situates our work in the extant literature on the impacts of son bias. In particular, our paper is the first to show that son bias affects HIV spread in Africa, and to document the nature of the mechanisms driving this effect. Section 3 presents our conceptual framework. The data and the empirical strategy are presented in Section 4. Section 5 presents the main findings and Section 6 presents the mechanisms. Section 7 concludes.

1.2 Contribution to the Literature

This paper is the first to establish a causal link between preferences for sons and HIV spread, and to study the possible mechanisms underlying this relationship. We view our work as contributing to the economics literature on the social and economic impacts of gender bias. Several studies have documented the impact of son bias on fertility choices (Jayachandran & Kuziemko (2011), Seidl (1995), Campbell & Campbell (1997), Clark (2000), Filmer et al. (2008)). These studies, which primarily, focus on Asia, find that son bias increases fertility. To our knowledge, our paper is the first to uncover a similar result in the context of sub-Saharan Africa. Other studies have investigated other outcomes. Mughal & Javed (2018) and L. Li & Wu (2011) show that having a larger number of sons positively affects a wife's

[6] Dahl & Moretti (2004) show that mothers of daughters are more likely to be divorced than mothers of sons. Similarly, L. Li & Wu (2011) and Mughal & Javed (2018) show that having a son increases a wife's bargaining power. Our data indeed shows that only widows who bore at least one son inherited from their late husband, while those who had only daughters saw their husbands' properties redistributed to his family members (figure 1-7).

[7] This result is close to the findings of Dahl & Moretti (2004) and Choi & Hwang (2015). These authors show that fathers are more likely to stay in the same house with their sons than with their daughters.

bargaining power in the household. Dahl & Moretti (2004) show that preferences for sons over daughters affect relationship stability, and that mothers who have only daughters have a greater probability of remaining single. Son bias has also been shown to lead to the abortion of female fetuses and the neglect and higher morbidity and mortality of girls in South-East Asia (Sen (1990), R. Pongou (2008), Anderson & Ray (2010), Milazzo (2018), Jayachandran (2017a)). It has also been shown that son bias in India leads mothers to breastfeed sons more than daughters (Jayachandran & Kuziemko (2011)).

The literature on son bias has generally focused on Asia, Europe and North America, with little attention paid to sub-Saharan Africa. It is generally agreed that the extent of son bias in Africa is very limited (Gangadharan & Maitre (1999), Morduch (2000), Mabeu & Pongou (2019)), in part because sibling sex composition in this region has not been found to have similar manifestations as in Asia.[8]. Using the same methodology as in the aforementioned studies and different measures of son bias, we uncover a positive effect of son bias on the likelihood of HIV infection, with this effect being visible primarily among women. Our mechanisms also uncover new findings on the impact of son bias. To our knowledge, we are the first to find that son bias affects sexual infidelity, increases the incentive to take on additional wives provided that the first does not have only girls, and increases the lifetime number of sexual partners, all of which affect HIV.

We also view our paper as contributing to the literature on the determinants of STIs. There is a wealth of literature on the determinants of HIV in sub-Saharan Africa. HIV in this region has been linked to poverty (Parkhurs (2009), Temah (2009)), commercial sex (Dupas (2005), Rao et al. (2003)) , age asymmetry between sexual partners (Duflo et al. (2015)), migration (Oster (2012), legal origins (Anderson (2018)), ethnic divisions (R. Pongou (2010)), configuration of sexual networks (R. Pongou & Serrano (2013), Kremer (1996)),slave trade (Bertocchia & Dimico (2019), "The Long-Term Effect of Demographic Shocks on the Evolution of Gender Roles: Evidence from the Transatlantic Slave Trade" (2014)), and ICT (Abbasi & Pongou (2022). By documenting a positive impact of son preferences on sexual behavior and HIV, our work adds to this literature. In addition, as already mentioned, we uncover several novel mechanisms that are interesting on their own rights, further enriching this literature.

[8] An exception is F. Arnold (1997), who finds a positive association between having only girls and polygamy in sub-Saharan Africa. However, our finding differs, especially when one focuses on first wives; a husband is less likely to take a second wife when his first wife has only girls. We uncover a positive effect of having only daughters on the likelihood of marrying a polygamous husband in the whole sample of married women. Together with the result uncovered for the subsample of first wives, this latter finding can be explained by the fact that mothers who have only daughters are more likely to find husbands who already have other wives. Such mothers therefore have limited prospects on the marriage market, as confirmed by our results on the effect of having only girls when unmarried on marriage timing.

1.3 Conceptual Framework: How Does Son Bias Cause HIV?

In this section, we present some theoretical arguments for why son bias may affect the likelihood of HIV infection in sub-Saharan Africa. We discuss five different channels: (1) unprotected sex through fertility choices; (2) extramarital relationships; (3) polygamy; (4) likelihood of getting married; and (5) lifetime number of sexual partners.[9] Our framework is summarized in Figure 1-8 below.

The literature that documents a link between son preference and family size argues that a couple will continue having children until it bears the desired number of sons. But child conception in the context of sub-Saharan Africa, where in-vitro fertilization is rare, implies having unprotected sexual relationships. This means that son bias is likely to increase HIV infection through increasing unprotected sex, especially in the presence of infidelity within a couple. We will test this mechanism by testing the effect of son bias on number of children. In fact, to the extent that the realization of a given sex is random (sex-selective abortion is non-existent in sub-Saharan African societies), son bias will lead to large family size, the latter reflecting the extent of unprotected sex.

The second channel that links son bias to HIV spread is sexual infidelity. The classical biological theory of sex determination among humans implies that the sex of a fetus is primarily determined by the male partner during conception. Each female child receives an X chromosome from each parent, while a son receives an X chromosome from his mother and a Y chromosome from his father (Stevens (1905), Wilson (1905), Sutton (1903), Henking (1891)). According to this XX-XY sex determination system, males carry the X and Y gametes while females carry only the X gametes, and a son is obtained by sperm cells that carry the Y chromosome. Owing to this theory, it is admitted that the father –or his sperm– determines the biological sex of the child. Therefore, a woman who had only daughters from her husband may consider bearing a son from a partner other than her spouse. In a context where paternity test is rare[10], a son obtained from a partner other than the husband will generally be attributed to the latter. The main incentive for a married woman to do so would be to stabilize her marriage and increase her bargaining power in the household. Indeed, studies show that a couple that has produced only daughters is less likely to be stable (Dahl & Moretti (2004)), and mothers who have more sons have greater bargaining power (Mughal & Javed (2018), L. Li & Wu (2011)).

[9]The last two channels can be viewed as related, because a woman who is unable to find a stable relationship is likely to have a higher number of sexual partners.

[10]It was only in 2014 that Congo obtained its first advanced equipment capable of analyzing DNA; ordered by the center of human genetics of the faculty of medicine of the university of Kinshasa.

The XX-XY sex determination theory implies that, Unlike wives, husbands have no real incentives to seek extramarital relationships to satisfy their quest for a son. In fact, since males determine the biological sex of a child, initiating an extramarital sexual relationship for a man who could only give daughters to his wife is likely to result in another female child. We will see below that the traditional theory of sex determination is incomplete because recent theories argue that female characteristics also play a significant role in determining whether a child is conceived male or female. However, we argue that even if an extramarital relationship could result in a son, in most African societies, such a son cannot become his father's heir (Mair (1971), Nors (1996)). It follows that we expect son bias to only lead to wife infidelity.

The third channel through which son bias may affect HIV in the African context is polygamy. Unlike the XX-XY sex determination theory, modern theories of sex determination argue that mothers play a role in the sex ratio of their offspring (W. H. James (1996), Wibbels et al. (1994), James M Cook (1995), W. James (1995), Crews (1996)). According to these theories, the offspring sex ratio is related to levels of hormones around the time of conception, and levels of parental hormones are themselves related to such factors as diet, occupations, stresses, illnesses, and exposure to certain environmental hazards.[11] It follows that, in societies where polygamy is practised, a man who had only daughters from his first wife may have an incentive to take a second wife in the hope that he will get a son (and a heir) from the latter. We therefore expect that mothers who have a high propensity to bear daughters are more likely to find themselves in a polygamous household. Polygamy is likely to increase HIV because it increases the size of sexual networks, which makes it possible for a random sexual infection to spread widely (R. Pongou & Serrano (2013)). Polygamy can also increase the likelihood of HIV infection, because if the second wife is infected before marriage, she will infect the first wife through their common husband.[12]

The fourth channel through which son bias may affect HIV is the difficult marriage market of women who bore daughters when unmarried. Studies show that a pregnant woman expecting a boy is more likely to be married with the father of her son at the delivery than

[11] For instance, it has been found that boys are more likely to be born to wealthier and more educated parents (Almond & Edlund (2007); W. James (1998); Trivers & Willard (1973)); to educated mothers (R. Pongou (2013b)); and to mothers with such medical conditions as type 2 diabetes (Moller et al. (1998), Paterson (1998)). Girls are more likely to be born to mothers with low gestational weight (Cagnacci et al. (2004)); to mothers with type 1 diabetes Rjasanowski et al. (1998)); and to aged parents (Lindahl-Jacobsen et al. (2000); Juntunen et al. (1997)). See R. Pongou (2013b) for a brief review of the related literature on parental factors affecting the offspring sex ratio.

[12] The polygamy channel developed above only focuses on first wives. However, this channel can also be extended to unmarried mothers. In fact, because women who bear a daughter when unmarried have a difficult marriage market (see below), such women may end up accepting a marriage offer from a man who already has other wives even if this is not her first choice.

if she expects a girl (Dahl & Moretti (2004), Mammen (2008), Lundberg & Rose (2003)). It follows that a woman who gave birth to a girl during her celibacy is more likely to remain single compared to her counterpart who bore a son. Similarly, the duration of celibacy is more likely to be longer for married women who gave birth to a girl before marriage compared to married women who bore a boy instead. It follows that a woman who had a daughter during her celibacy is more likely to be exposed to a larger number of uncommitted relationships than her counterpart who had a boy, leading to a greater likelihood of HIV infection.

The fifth channel through which son bias may affect HIV is by increasing the lifetime number of partners. This channel is partly a corollary of all the channels discussed above, except for the first channel. However, it can also be viewed as deriving from the notion that son bias also increases the likelihood of divorce in a couple that has produced only girls (Dahl & Moretti (2004)). Divorce in turns leads to new relationships, increasing both the lifetime number of sexual partners and HIV.

1.4 Data and Empirical Strategy

In this section, is described the data used for the analysis in this paper as well as our empirical strategy.

1.4.1 Data Description

This chapter exploits 153 Demographic and Health Surveys (DHS) from 37 sub-Saharan African countries. These surveys were conducted between 1999 to 2017 (table A.1). DHS are representative at the national and sub-national level, and most variables are comparable across countries and over time. Each survey uses a two-stage probabilistic sampling stratification, selecting clusters (census enumeration zones) in the first stage and households in the second stage. All household members are surveyed for general socioeconomic and demographic questions. In each household, information is collected for selected women and men on a wide range of questions concerning health behavior, fertility preferences and behavior, in-depth socioeconomic and demographic characteristics, and health status for several health outcomes (table 1-1). The DHS mostly targets women, but in many countries, some surveys also include men who are coupled with some of the interviewed women. DHS files include single women, but most DHS phases do not have single men in the sample. For this reason, when we analyze how the effects son bias differ for men and women, we restrict our samples to married women and men.

In particular, for relatively recent surveys, HIV blood tests are conducted for women aged between 15 and 49 and men aged between 15 and 59.[13] Only 45 surveys from 27 countries have information on HIV test results, yielding a sample of 201,495 observations for women and 63,070 observations for men. The next section presents four different measures of son bias and sibling sex composition used in our analysis. We merge our dataset with data from the Murdock's Ethnographic Atlas map (Murdock (1959)) using geographic references provided with the DHS files. Indeed, using the GIS software, we import the geospatial file of Murdock and the one containing the geographical characteristics of the DHS. We extract the overlapping portions of features in the DHS and the Murdock layers. We compute their intersection layer containing the attributes of the overlapping layer from both the DHS and the Murdock files, using the intersect command. This computation caused us to lose about 9 percent of our sample, because Murdock's data does not contain the characteristics for all ethnicities in the DHS.

The Murdock's ethnographic Atlas map is a commonly used anthropological database in social sciences. It provides information on the characteristics of more than a thousand ethnic groups.

Measuring Son Bias

Our measures of son bias partly follow from the economics literature on this topic. In this literature, the detection of son bias generally relies on testing the effect of sibling sex composition on some outcomes of interest such as fertility decisions (Basu & Jong (2010), Altindag (2016), Rai et al. (2014), Kugler & Kumar (2017), Bedard & Deschenes (2003)), breastfeeding (Jayachandran & Kuziemko (2011)), allocation of household resources (L. Li & Wu (2011)), and sex differences in child health and survival (Jayachandran & Pande (2017), Song & Burgard (2008)). We therefore rely ourselves on existing studies showing that parents' reaction to the realization of their offspring sex composition makes it possible to determine whether they present a bias or not; for example, given the offspring sex composition, parents will demand for additional children only if they are gender biased. We use four different variables, including two variables that measure sibling sex composition, and two variables that combine sibling sex composition and son preferences. Two of the four measures are inspired by the literature, and the two others are new.

The sex of the first child. The sex of the first child is a variable commonly used in works related to son bias (Dahl & Moretti (2004), Basu & Jong (2010), Altindag (2016), Rai et al. (2014), Kugler & Kumar (2017), Bedard & Deschenes (2003)). Dahl & Moretti (2004) use a binary indicator of whether the first child is a girl to demonstrate how offspring sex

[13]Some most recent surveys also have information on the HIV status of children.

composition affects transition to marriage and divorce in the United States. They show that mothers with a first child girl are less likely to be married and more likely to be divorced. This clearly demonstrates that fathers have a favorable bias for sons. Along the same line of research, Kugler & Kumar (2017) use the sex of the first child to prove that in India, parents with a female firstborn have larger families. Similarly, Basu & Jong (2010) find that daughters have a greater number of siblings and are usually among the first in the birth order.

Having only daughters. Sibling sex composition is also measured by the percentage of daughters among the offspring, a particular case being when all children are female (Garg & Morduch (1998), Jayachandran & Pande (2017), Butcher & Case (1994)). Using data from India, Jayachandran & Pande (2017) find that the demand for an additional child is constantly positive for biased couples having only daughters. It follows that biased parents with only daughters have larger family size, therefore allocating fewer resources to each child (Butcher & Case (1994)). Garg & Morduch (1998) use a similar approach to demonstrate that girls who have no brothers are on average 25 to 40 percent healthier than those who have at least one brother. Thus, the rivalry in the allocation of resources to children is favorable for children having only sisters. Following these studies, we will also measure sibling sex composition by a binary indicator for whether all children are female.

The natural (or biological) propensity to have more daughters than sons. We construct this variable by regressing a variable that indicates whether an individual has more daughters than sons on the total number of children ever born. The residuals of this regression measure the natural or biological propensity to have more daughters than sons. In order to avoid assigning the proportion of girls observed within countries to women without children -which could bias our results-, only women who have at least one child are included in the analyses. Analyses are done at the individual level. We use the mothers' birth history to compute their natural propensity to bear daughters. For men, we use the variables providing their total number of sons and daughters at home and elsewhere from the MR files to compute their individual level biological propensity to procreate daughters. Here too, only men who have at least one child are included in the analyses.

By controlling for the number of children, this variable accounts for fertility preferences and for son bias resulting in more children. Mothers with a greater propensity to have more daughters than sons may be more likely to be neglected by their husbands, which could lead to infidelity, even if this behavior does not reflect an attempt to have a boy. This will in turn lead to a greater likelihood of HIV infection.

Realized percentage of sons smaller than the ideal percentage. This variable is a binary indicator for whether the realized percentage of sons among the offspring is smaller than the desired (or ideal) percentage; computed at the individual level as well, for respondent with at least a child. It is therefore a combination of the offspring sex composition and parental preferences. It captures perfectly the bias for sons since a biased parent whose percentage of sons is smaller than his/her ideally desired percentage would exhibit a positive demand for an additional child. We construct this variable based on information on the realized and desired number of sons and daughters at the survey.

Summary Statistics

The summary statistics of the main variables used in our analysis are presented in Table (1-1) We note that there are more women than men in our sample because the DHS focuses mainly on women. In our sample, an average woman is 31.8 years old, and an average man is 31.4 years old. About 93.8 percent of women have ever been married, and among married individuals, 52.8 percent of women are involved in a polygamous relationship. Educational level is relatively low; 34.3 percent of men versus 34.1 percent of women have no education, and 36.6 percent of women versus 34.3 percent of men have primary-level education. We also note that 80 percent of men versus 60 percent of women want more sons than daughters, indicating that son bias is very high in our sample. HIV prevalence is 5.3 percent for men versus 6.1 percent for women.

Regarding the variables measuring sibling sex composition and son bias, we note that about 49.2 percent of firstborns are girls (for either men and women), and the percentage of daughters among children born to an average individual is 47.5 percent. Over 73.3 percent of mothers have at least as many daughters as sons. Also, the average biological propensity of mothers to bear more daughters than sons is -0.04 percent versus 0.04 for fathers[14]. Indeed, 26.93 percent of mothers versus 36.91 percent of fathers have more daughters than sons. We also note that 18.87 percent of women have only daughters versus 17.6 for fathers. Finally, about 46.6 percent of mothers have a son percentage smaller than what their ideal percentage versus 43.2 percent of fathers.

[14]Note that the discrepancy in this statistic between men and women may be explained by several factors, including the fact that some men have multiple wives.

1.4.2 Identification Strategy

Studies analyzing the causal effects of son bias generally assume that the realization of child sex is purely random in the absence of sex-selective abortion. In the African context, it is shown that sex-selective abortion is non-existent (Carrera (2011), Cumber & Nchanji (2017)).[15] Following the literature, we therefore estimate the following model using ordinary least squares with fixed effects:

$$HIV_{i,c} = \beta_0 + \beta_1 bias_i + \beta_2 Educ_i + \beta_3 age_i + \beta_4 reside_i + \beta_5 wealth_i + \beta_6 religion_i + \gamma_{h_c} + \theta_t + \varepsilon_i \tag{1.1}$$

In this equation, $HIV_{i,c}$ measures the HIV status of individual i, residing in country c; it takes the value of 1 if i is infected with HIV and 0 if i is not. The variable $bias_i$ measures offspring sex composition or son bias for individual i, using any of the four measures described in section 4.1.1. We control for individual i's education level ($Education_i$), age (age_i), type of place of residence ($reside$), wealth index ($wealth$), religion ($religion$), year fixed effect (θ_t), and for ethnic homeland by country fixed effect (γ_{h_c}). Finally, ε_i is a random error term assumed to be uncorrelated with son bias given the controls.[16] All analyses in this paper are entirely done at the individual level using DHS weights.

1.5 Findings

In this section, we first present the causal average effect of sibling sex composition and son bias on HIV (section 1.5.1). This is followed by an analysis of heterogeneity in these effects by age at first birth, by education, by ethnic characteristics and development levels.

1.5.1 Average Effect of Son Bias on HIV Spread

The estimated effects of sibling sex composition and son bias on HIV status are reported in table (1-2); panel A and C reports the results for women, and panel B reports the results for men. Each table has four columns, with each showing the effect of a specific measure of son bias (or sibling sex composition) on HIV status. All columns include baseline controls including age, place of residence, education, wealth, religion, year and ethnic homeland by

[15] These studies argue that medical equipment to monitor pregnancies are lacking in most sub-Saharan African settings. Even in most urban areas where such equipment is likely to be found, there is no evidence of distorted population sex ratio at birth in the data.

[16] These additional controls do not appear to affect the estimates of the effect of son bias on HIV, which also confirms that our results are unlikely to be driven by unobserved variables. In particular, almost all of the preconception factors thought to affect the sex of a fetus are unlikely to affect sexual behavior and HIV.

country fixed effects. Panel A presents the results obtained on the sample of women who responded to the HIV surveys, and panel C is restricted to women who are in union with the men in panel B. This allows to eliminate sampling bias, and strengthen the robustness of our results.

In column 1 of table (1-2), we find that having a female firstborn gives a mother a likelihood of being infected with HIV of 0.33 percentage points more higher in comparison to a mother whom the firstborn is a boy. Column 2 shows that increase in the natural propensity to have more daughters than sons also has a large effect. A one-unit increase in this propensity increases a mother's likelihood of HIV infection by 0.45 percentage points. Similarly, having only daughters increases a mother's likelihood of getting infected by HIV by 0.6 percentage points (column 3). Furthermore, for mothers whose realized percentage of sons is smaller than their ideal, the risk of infection is 0.21 percentage points higher. For parents of daughters, in the presence of a favorable bias for sons, the odds of mothers being HIV infected is positively significant at conventional levels, and are large, especially given the prevalence of HIV in our sample, which is over 6 percent.

These results all show that bearing daughters increases the risk of HIV infection for mothers.Given the binary nature of the outcome, we also estimate the effect of son bias using other functional specifications such as Logit and probit[17], and we find qualitatively similar results. These clearly demonstrate one manifestation of son bias in African societies. Indeed, bearing daughters in a social context where a greater value is placed on having sons is likely to encourage mothers to opt for riskier solutions that could allow them to have a son. These options will be explored later when we analyze the mechanisms.

Panel B of the same table presents the results of the same estimations on the sample of fathers. We find opposite effect among fathers. Interestingly, the effects of our four measures of sibling sex composition on son bias are negative, although they are no statistically significant for the indicator of whether the realized son percentage is smaller that the ideally desired one and the indicator for whether the father has only daughters.

The results in panel C, obtained from similar regressions on the sample of women in a marital union with the fathers in panel B reinforce the validity of our conclusions. Indeed we find the effects of having mostly daughters and of son bias are positive and statistically significant for these women, while the effects are negative for their husbands.

Moreover, when comparing the prevalence of HIV in each of the four groups, one can see variation of about 1 point among women, showing that the offspring gender composition of a mother, given the preference for sons, has a substantial impact on her serological status.

[17] See Figure 1-6 for the graphical representation of the relationship between a mother's natural propensity to have more daughters than sons and HIV.

However, by carrying out the same comparison among fathers, we rather see an absence of variation in the percentage of infected fathers in relation to the composition of the sexes of their offspring. So clearly, preference for sons does not increase men's HIV, and not even for fathers of daughters. This result is interesting because our data reveal that men have a more pronounced preference for sons than women. The findings therefore seem to indicate that women are more likely than men to seek multiple partners in response to son bias, which is consistent with the classical XX-XY sex determination theory. We explore and validate this explanation in the mechanism section.

1.5.2 Heterogeneous Effects of Son Bias on HIV Spread: Individuals' Characteristics

In this section, we estimate heterogeneous effects of sibling sex composition on HIV status by level of son bias. We also analyze heterogeneity in the effect of sibling sex composition and son bias on HIV status by education level and by age at first birth. These analyses are important to start understanding the mechanisms through which sibling sex composition and son bias affect HIV in sub-Saharan Africa.

Is It Really Son Bias?

In order to further investigate whether son bias is driving the effect of the offspring sex composition on female HIV status, we study heterogeneity in this effect by level of son bias. In our sample, some individuals are biased toward sons in the sense of desiring more sons than daughters, and some others are biased toward daughters. Given the randomness in the determination of child sex, these two groups of mothers are comparable in terms of their offspring sex composition. There is also is a great deal of heterogeneity in offspring sex composition among these mothers, which makes it possible to investigating the HIV effect of sibling sex composition in each group. Following equation (1), we regress the binary indicator for HIV on each of the variables measuring sibling sex composition, controlling for the other variables as in Table (1-2). The results of these regressions are shown in table (1-6).

We find that almost all of the variables used to measure the offspring sex composition and son bias have a null effect on the likelihood of HIV infection among daughter-biased mothers. Only the binary indicator of if the individual has only daughters has a 5 percent level of significance and a positive effect on HIV in this subsample; a daughter bias mother having only daughters is 1.44 percent more likely to be HIV infected, in comparison to her counterpart who has at least a son. We do the same analyses for son-biased mothers. We find that each of the variables used to measure the offspring sex composition and son bias has a positive and statistically significant effect on HIV. The results are presented in the second panel of Table (1-6). We note that, while the probability of mothers being infected increases by 0.04 points among those with a bias for daughters, that of mothers in the second group increases by 0.06 points, with their propensity to have more daughters than sons (column (2)). It is also interesting to note that HIV prevalence is relatively higher among mothers with son bias. This observation therefore supports the thesis that son bias is a factor that favours the persistence of HIV prevalence in sub-Saharan Africa.

In general, the findings suggest that it is a preference for boys that is driving the positive effect of offspring sex composition on HIV among women, since this effect mainly exists among daughter-biased mothers.

We use another specification to check the robustness of our analysis. In this specification, we regress HIV status on the variable measuring the degree of son bias as proposed by Jayachandran (2017b)[18] across four subsamples, all characterized by the gender composition of their offspring. This measure is the residual of the regression of a binary variable for whether one prefers more sons than daughters on the total number of children[19], controlling for all the baseline variables in equation (1.1).

The results of this analysis are presented in Table (1-7). We find that, in comparison to parents of mostly sons, parents of mostly daughters are more likely to be infected under the effect of son bias. Indeed, Column (1) shows that son bias yields an odd to be infected 0.07 higher for mothers whose first child is a girl compared to mothers whose first child is a son. Similarly, son bias marginally increases the odds of mothers with a greater natural propensity to bear daughters by 0.07 points. This finding is consistent across all four subsamples. In column 3, mothers with a son bias who have daughters only are 0.07 percentage points more likely to be infected than biased mothers having at least one son. Thus, gender biased parents having mostly daughters are on average 0.07 percent more likely to be HIV

[18] By observing the numbers in the DHS database, one can notice that people who want an even number of total desired children may present an undeniable preference for a given sex. For example, an individual who wants 4 children will strictly desire one sex more than the other or else, will have equal preferences for each of the two genders. However, when the number of children desired is an odd number, it is obvious that it will be more difficult to appreciate a strict preference for a given gender when the difference between the number of daughters wanted and that of sons wanted is equal to one. To determine the level of preference for sons, Jayachandran (2017b) proposes to regress the number of sons desired on the total number of children desired and to keep the residual of the regression as the real son preference indicator. We also use this definition of son bias to assess the impact of son preference on HIV.

[19] The idea is that the measurement should account for fertility behavior and for the fact that the number of children desired by an individual might be an odd number. If the desired number of children is odd, then it is clear that the desired numbers of boys and girls will be different, although this does not necessarily reflect a preference for a particular sex.

infected than parents of sons. Moreover, the higher prevalence of HIV among mothers with a higher propensity to bear daughters is consistent with the hypothesis that son bias is a factor that increases the risk of HIV infection among parents of daughters.

Heterogeneous Effects by Education Level

We have argued that the XX-XY sex determination theory might explain why son bias causes women to be unfaithful, unlike men. However, in most countries, this theory is taught only in secondary school. It follows that educated individuals are more aware of this theory than their counterparts who are uneducated. This implies that if the XX-XY sex determination theory is an explanation for why son bias causes female infidelity, the effect of son bias on infidelity and HIV status should be more pronounced for educated mothers and less pronounced for educated fathers. We test this hypothesis and report the results in table (1-3) for mothers and in table (1-4) for fathers.

In table (1-3), we find that the effect of having daughters on HIV is much greater for educated mothers (or mothers with at least a secondary level education). This is consistent across the different measures of son bias and sibling sex composition we analyze in this paper. For example, the effect of having a female firstborn on the likelihood of HIV infection is about 0.3 percent more for uneducated mothers compared to their educated counterparts vs 0.47 among more educated mothers. Similarly, the effect of a one-unit increase in the natural propensity to bear more daughters than sons on HIV infection is 0.90 percentage points greater for educated mothers compared to 0.30 for uneducated mothers. Also, the effect of having relatively fewer sons than desired on HIV is 0.56 positive and statistically significant for educated mothers and only 0.12 and statistically null among less educated mothers. The patterns are opposite for fathers (table (1-4)). The effects of son bias on HIV are all negative and statistically significant for more educated fathers compared to their educated counterparts. For all our predictors, the effects are negative and statistically significant for educated fathers, but presents greater coefficients among the less educated fathers. But for most measures, the differential effect is not statistically significant, except for few cases.

These findings largely validate our hypothesis that if the XX-XY sex determination theory is able to shed light on our main results, then it should be more valid for educated individuals than for their less educated counterparts. For women, this means that sibling sex composition and son bias should increase HIV spread more among educated mothers and less among educated fathers, which is exactly what we find.

Heterogeneous Effects by Age at First Birth

The effect of sibling sex composition and son bias on HIV might depend on the fertility horizon of an individual. A woman who started having children early (or has a long fertility clock) might be more optimistic about having a boy even if her first children are girls than a similar woman with a reduced fertility clock. If so, then the effect of son bias on HIV should be larger for mothers who started having children later. But the effect of son bias on HIV should be non-linear with respect to age at first birth, as women who have a very short fertility clock might be more concerned about just having a child, regardless of the gender.

In order to test this hypothesis, we divide our sample into three subsamples representing age at first birth: 15-24; 25-39; 40 and over. The estimations are done for each group, and the results are presented in table (1-5) for mothers and in table (A.2) for fathers. We find that the effect of son bias on HIV is larger for mothers who started having children when they were between 25 and 39 years old. The effect of having only female children is 0.49 percentage points for mothers whose age at first birth lies in the interval 15-24 and statistically nil for mothers who had their first childbirth at 40 and over. In panel B, mothers who have only daughters are 2 percentage points more likely to be infected than those who have at least one son. Having only daughters has a positive and statistically significant effect for the former group, although this effect is much larger for women who started having children at age 25-39. These effects are generally not present among fathers (table (A.2)), reflecting the null average effect. Among mothers, the effect of son bias on HIV with respect to age at first birth is indeed non-linear, being in general larger for women who started having children at age 25-39.

The findings are consistent with the conjecture that women who start having children at a young age are more optimistic about having a boy even if they started having daughters. However, at age 40 or older, a woman is more concerned about just having at least one child.

Heterogeneous Effects by Development Level

Figures (1-10) and (1-11) show that there is a positive correlation between economic development and son preference in sub-Saharan Africa. While one would expect the development to be accompanied by a reduction in discrimination against women, these very interesting statistics reveal the opposite. Hence the dual need for development policies in sub-Saharan Africa to pay more attention to issues of discrimination suffered by women, which is still very predominant in this part of the world.

Starting from the fact that it is a proven statistic—the positive relationship between economic development and preference for sons in Africa—we hypothesize that son bias will have a greater effect on the prevalence of HIV in sub-Saharan Africa among more developed nations in comparison to less developed countries.

We verify the validity of this proposition by estimating equation (1.1) on two samples. The first sample is that of women living in the more developed countries of sub-Saharan Africa. The second is that of women living in the least developed countries of this same region. In fact, we have grouped countries according to the level of economic development as classified by their 2019 GDP.[20]

The results of these estimates are presented in table (1-12). Son bias is a factor that increases the likelihood of being HIV-positive in both groups. However, the coefficients are more than twice as high for women living in more developed countries of sub-Saharan Africa (Panel A). Indeed, while women in less developed countries who had borne only daughters are 0.32 percentage points more likely to be infected, those in more developed countries are 1.49 percentage points more likely to be infected. This result not only validates our main hypothesis, but it also reinforces the thesis that development is positively correlated with son preference.

Logically, one would expect economic development to be followed by a reduction in the prevalence of discrimination against the female gender because of a greater participation of females in socio-economic and political activities. That we do not find this in the analysis is intriguing and requires additional investigation.

1.5.3 Heterogeneous Effects of Son Bias on HIV Spread: Ethnic Groups' Characteristics

We use a different specification to check the robustness of our analysis. In this specification, we regress HIV status on a variable measuring sibling sex composition, controlling for all the baseline variables in equation (1.1) as well as for the type of place of residence, wealth, religion, and the ethnic homeland fixed effects. The results of this robustness analysis are presented in table(1-8). We also evaluate the validity of our results on different samples based on the type of kinship, the tribes' marriage modes, polygamy practices.

1.5.4 Patrilineal/Matrilineal kinship, Son Bias, and HIV

Table (1-8) contains the results of the estimation of equation (1.1) on the sample of mothers in patrilineal and matrilieal societies, respectively. The results suggest that son bias affects

[20] According to this classification by level of 2019 GDP, the nine countries with the highest levels of GDP per capita are, in descending order, Gabon, South Africa, Namibia, Angola, Nigeria, Ghana, Kenya, Zambia, and Lesotho. The other countries in our sample are considered to be less developed.

has a greater impact on the risk of HIV infection in patrilineal societies than in matrilineal societies.

The findings are consistent with those of our main results. We find that son bias increases the likelihood of being HIV infected. In general, the significance of the coefficients is similar to that obtained from the main regressions (table(1-2))

Panels A and B display the impact of son preference on HIV within patrilineal and matrilineal communities, respectively. Within patrilineal communities, preference for sons increases the likelihood of mothers being HIV-positive. As the propensity of mothers of patrilineal tribes to have daughters increases, their probability of being infected also marginally increases by 0.46 percentage points (column (2)). Similarly, in patrilineal tribes, mothers who have only daughters are 0.66 percentage points more likely to be HIV positive compared to their counterparts who have at least one son. All the results in panel B support the hypothesis that having daughters in communities with son bias increases the risk of HIV infection.

However, within matrilineal societies, having daughters does not significantly affect the propensity to be HIV positive. On the contrary, we find in column (2) that in matrilineal communities, as the propensity to have more daughters increases, the probability of being infected decreases by 0.69 percentage points.

Figure (4) clearly illustrates this result. We can see that the probability of being infected increases with the percentage of daughters in patrilineal communities, while it decreases with the propensity to have daughters in matrilineal tribes. These results are consistent with the hypothesis that son bias is a driver of the persistence of AIDS.

1.5.5 Polygamy, Son Bias, and HIV

We also study heterogeneous effects of son bias between tribes where polygamy is predominant against those where marriages are mostly on a monogamous regime. Results are displayed in table(1-9). For both groups, we find positive and highly significant effects of son bias on the probability of being infected with HIV. However, the coefficients are considerably lower within monogamous tribes. Women living in communities where polygamy is predominant, when their first child is a girl, are 0.40 percentage points more likely to be HIV infected compared to others whose first child is a son (versus a non significant effect of 0.20 percent in monogamous tribes).

The risks of HIV infection for those with a strong tendency to produce girls are therefore generally higher within polygamous tribes compared to monogamous tribes. As we show in the next section, while son bias induces fathers of daughters to take other wives who could give them sons, it increases mothers of daughters' incentive to engage in infidelity. In the first case, the husbands are undeniably the main vector for the spread of the virus. The positively significant coefficients within polygamous tribes simply reinforce the conjecture that men are physiologically less vulnerable to sexual infections than women. Women in a monogamous tribe (where the man cannot take other wives), who produce only daughters have a great incentive to engage infidelity. In this other case, mothers appear as the main vectors of spread of potential sexual infections. Because of their higher physiological vulnerability to this type of infection, by relying upon infidelity to bear a son, they will necessarily have a greater of risk of HIV infection. This is all the more true as in tribes where polygamy is part of the customs, we have higher HIV prevalence, reflecting the fact that women are more vulnerable to HIV than men

All these results reinforce the idea that preference for sons is a factor allowing the prevalence of HIV when there is a greater tendency to produce daughters.

1.5.6 Bride Price, Son Bias, and HIV

We estimate equation (1.1) for tribes where the bride price is compulsory and in tribes where it symbolic. In table (1-10), we find that in tribes where the bride price is optional or symbolic, having daughters has no significant impact on the probability of being infected, with coefficients negative and statistically not significant (Panel A). However, in Panel B of this same table, we find that in communities where the bride price is mandatory, having daughters increases the likelihood of being HIV positive. Indeed, in these communities, as a woman's propensity to have more daughters increases, her likelihood to be infected marginally increased by 0.52 percentage points. In the other group, the coefficients are negative and not statistically significant.

As we can see in table (A.3), tribes where bride price is practiced are essentially patrilineal. Therefore, the significant results are consistent with the findings according to which preference for sons is a factor determining the prevalence of HIV in patrilineal tribes. It also reflects, however, the fact that the mandatory nature of the bride price or the wealth paid to the bride's family, in a context where son bias is fundamentally important, increases the social cost associated with the fact of having a great propensity to bear daughters. This would encourage women in such societies to take actions likely to increase their risk of HIV infection. Moreover, this social cost, necessarily lower in tribes where the bride price is optional or symbolic, favors risky behavior less among mothers with a strong tendency to have daughters. This, therefore, would explain the lack of statistical significance in the first panel of table (1-10). Table (1-11) confirms this. Indeed, we compare the impact of son bias in patrilineal tribes where bride price is mandatory, to that of matrilineal tribes where the bride price is optional or symbolic. In the first group, we have coefficients that are higher than those of the second group, but also, only those of the first group are statistically different from zero. This reinforces the proposition that the social cost associated with producing only daughters is necessarily higher in tribes where the bride price is an obligation.

The result illustrated in Figure (1-9) is consistent with the latter. We use the bias level indicator as defined by Jayachandran (2017b) to show the relation between son bias and this effect differs by bride price practice. The graphs show that in tribes where the bride price or wealth is a token, the probability of being infected decreases with the bias towards sons, while this probability increases as son bias increases within communities where the bride price is mandatory.

This finding obviously reflects the social cost imputed to mothers in one or other of the tribes. This social cost, associated with the fact of having a great propensity to bear daughters within patrilineal communities (in essence characterized by a strong preference for sons) is greater in tribes where bride price is mandatory.; equivalently, the value given to a son is strictly greater than the bride price received by a daughter's family.

1.6 Direct Mechanisms

In this section, we analyze several direct mechanisms that may explain our main finding that son bias increases the likelihood of getting infected with HIV among mothers, but not among fathers. Following our conceptual framework (section 3) represented by figure (1-8), we test five distinct channels: unprotected sex as reflected by family size; extramarital relationships; polygamy; marriage prospect of single mothers; and lifetime number of sexual partners. These channels are tested for women mostly, except for sexual infidelity which is also tested for men.

1.6.1 Son Bias and Unprotected Sex

Reproductive choices may be one channel through which son bias increases the risk of HIV infection. As already noted, parents want more sons than daughters in our sample. Therefore, when the offspring sex composition is unsatisfactory, it is possible that they will demand for additional children, which implies engaging in unprotected sex in a context such that of sub-Saharan African societies where in-vitro fertilization is almost absent. This in turn will increase the likelihood of HIV infection, especially if one partner has other sexual partners. We test this hypothesis by estimating the effect of son bias on the number of children born to a mother. We test this hypothesis on the sample of all mothers, the subsample of mothers who are son-biased, and the subsample of daughter-biased mothers. We estimate equation

1, where the dependent variable is now the number of children. The results are presented in table (1-13). Regressions are carried out both on the samples of women who responded to the HIV surveys and those who responded to the surveys of women. We find that our results show no sampling bias, as they are robust regardless of the sample used.

Panel A presents the results for the sample of all mothers. We find that having a female firstborn has a positive and statistically significant effect on the number of children born to a mother. The effect is consistent and statistically significant across different subsamples of mothers. Having a female firstborn increases the number of children by 2.72 percentage points in the sample of all mothers, by 3.24 in the sample of mothers who have ever been married, and by around 3.02 in the sample of single mothers. Also, the effect is the largest in the sample of mothers whose first childbirth occurred during celibacy (4.95). These coefficients are indeed very important in relation to the average number of children in each sample, which varies between 2 and 4.

The results for the subsample of son-biased mothers (in Panel B) contrast those obtained for the subsample of daughter-biased mothers (in Panel C). Having a female firstborn has a 1.87 positively significant effect on the number of children for son-biased mothers but a lower and non statistically significant effect for daughter-biased mothers. The findings clearly indicate that son bias is driving the positive effect of having a female firstborn on number of children. This finding is consistent across all the four subsamples. The demand for additional children is higher among son-biased mothers in comparison to daughter-biased mothers across all the four subsamples.

Remark that we use only one of the four indicators for son bias and sibling sex composition for the analysis. The reason is that the computation of the other indicators already relies on the number of children, and using them would therefore create an obvious endogeneity problem. However, we investigate the effect of these indicators on the "desire" to have additional children in the subsample of mothers who already have more children than ideally desired. The results are presented in table (1-14). We find that having a female firstborn, or having a greater propensity to bear daughters, or having only daughters, or having a smaller number of sons than the ideal number increases the desire for more children, even though the parents have already exceeded the ideally desired number of children.

These findings, which clearly show that son bias increases family size and the desire to have more children even among mothers who already have more than their ideal number, imply that son bias increases the likelihood of unprotected sex over the fertility horizon. This in turn is likely to increase the risk of HIV infection.

1.6.2 Son Bias and Extramarital Infidelity

The second channel we investigate is extramarital relationships. As argued in our conceptual framework based on the classical XX-XY theory of sex determination, son bias may provide an incentive for sexual infidelity among women, but not among men. We test this hypothesis for women and men who have been married or living with their partner for at least one year. We estimate equation 1, where the dependent variable is a binary indicator for whether an individual has had an extra-marital relationship within the past 12 months preceding the survey.

We also want to identify which of the biases between that of mothers and that of their husbands is more prone to induce the women into infidelity. To do this, we perform the same regressions on the samples of women married to the men interviewed and who present a positive bias for sons. Our results are not only consistent with the idea that the presence of son bias increases mothers' propensity to commit infidelity, but more interestingly, their husband's bias has a greater effect on their propensity to cheat than their own bias. At the same time, husbands are not affected by their own bias.

The results are presented in table (1-15) and (1-16) for women and in table (1-17) and (1-18) for men. Findings are shown for the whole sample of men and women, and we also investigate heterogeneous effects by marriage duration.

In Panel A of table (1-15), we show results for the whole sample of married women. We find that when the number of daughters is greater than ideally desired, positively affects infidelity among women. For instance, the probability of marital infidelity is 5.89 percentage points higher among women who have fewer sons than desired compared to other women. However, these effects vary significantly by marriage duration. There is little effect of son bias on marital infidelity for women who have been married for at most five years. In fact the effects are almost all negative. Only one predictor (having fewer sons than the ideal number) positively affects infidelity among these women. The effects are stronger and have the expected sign among women who have been married for a longer period (see Panel c and panel A and B of table (1-16)). These results seem to suggest that women who have been married for less than 5 years are more optimistic about having a son with their partner and that pessimism grows with marriage duration, pushing wives to seek extramarital relationships to achieve their desire for a son when they perceive that their husbands may be unable to give them one.

As for husbands, we find no effect of son bias on their propensity to be unfaithful (table (1-17)). The sign even appears to be negative for most predictors for the whole sample (Panel A) and for marriage duration up to 10 years (see Panels B and C), although not generally statistically significant. These findings are consistent with the XX-XY theory of sex determination according to which, within a couple, the man is the one who determines the sex of the child. Modern generalizations of this classical theory also assign a role to women. However, even in this case, men will still have no incentives to have extramarital relationships to have a son, given that a son born outside of a recognized or official relationship will not inherit from his father in most African societies (Osaranen (2008), Mair (1971), Nors (1996)). These findings also seem to suggest that the strong preference for sons in sub-Saharan African societies only leads women to unfaithfulness, perhaps in the hope that having a son, even from an extramarital partner, will contribute to stabilizing their marriage. On the right of these tables, we can see that the coefficients are larger, with a greater significance for women whose husbands have a positive bias for sons. This leads us to conclude that men's bias has a greater impact on the infidelity of their wives.

In the next section, we explore polygamy, which is more consistent with modern generalizations of the XX-XY theory of sex determination.

1.6.3 Son Bias and the Marriage Market

In this section we analyse the effects of a preference for sons on marriage outcomes. The marriage outcomes we analyze are polygamy and the likelihood of finding a husband for women who had their first child before marriage.

Son Bias and Polygamy

We analyze the effect of the offspring sex composition on the likelihood of marrying a polygamous husband. In order to control for selection into marriage, the analysis is restricted to married women. We also study heterogeneous effects by age at survey.

We estimate equation 1, where the dependent variable is a binary indicator for whether a woman is married to a polygamous man. To clearly show that men's bias is a factor that drives them into polygamy, we perform the same regressions on the sample of women whose spouses have a positive bias for sons. The results are consistent with the fact that men prefer to have a legitimate son who can be their heir rather than opting for infidelity, which would allow them to have an illegitimate son who cannot inherit from them. Therefore, polygamy is a rational choice for fathers of daughters rather than infidelity. The results are presented in table (1-19). We find that all the four variables used to measure the offspring sex composition positively affect the likelihood of marrying a man who has other wives (see Panel A of table (1-19)). The effects are statistically significant at conventional levels, except for the effect of having a female firstborn. Panels B-D report the results of the estimation for women aged 15-29 years old, 30-44 years old, and 45 years old or older, respectively. We see that the findings are consistent across these different age groups, although they generally tend to be more pronounced for older women. In panel B, having more daughters than desired increases the probability of marrying a polygamous man by 5.64 percentage points for mothers aged 15 to 29 years old, while for the older group, the coefficients are highly significant and generally higher. Hence, it is not only the bias that favors polygamy, but the propensity to procreate daughters in an environment where there is an undeniable preference for sons.

There are two plausible interpretations of the results reported in table (1-19). One interpretation is that women who are more likely to bear daughters have limited marriage prospects, especially when they have children before marriage. They end up marrying men who already have other wives, which may not be their first choice. Another interpretation is that husbands of women who do not produce the desired number of sons end up taking additional wives to achieve this goal. The results on the right side of the table effectively show that men who have a bias for sons have a positive probability of being polygamists when their offspring consists mainly of daughters. We explore with more precision this latter explanation, by focusing on first wives. The results are reported in table (1-20).

The findings of table (1-20) show that a higher (natural) propensity to bear more daughters than sons predicts the likelihood of a husband with positive bias, taking a second wife. With a 13.2 percent likelihood to take a second wife when his wife bore only daughters. This findings indicate that son bias husbands prefer taking additional wives when the first wife has a non-satisfactory number of sons, rather than opting for infidelity.. This result also allows us to conclude that the positive effect of having only daughters on marrying a polygamous man found for the whole sample of married women is driven by fathers' bias toward sons.

Son Bias and Celibacy

We hypothesize that bearing daughters when single could increase the likelihood of a woman to remain single or to experience marriage delay. Although remaining single or experiencing marriage delay does not by itself cause HIV, it may lead to an increased number of sexual partners during the lifetime, which would positively affect the likelihood of getting infected with HIV. We conduct this analysis on the subsample of women who were at least 45 years old at the survey and who had their first child during celibacy. Restricting the analysis to these ages ensures that women have been sufficiently exposed to the marriage market; also having a child after the age of 45 is harder, which decreases the probability of finding a husband after this age.

We estimate equation (1.1) where the dependent variable is now a binary indicator for whether a woman has never been married. The results are reported in table (1-21). Three of the variables measuring the offspring sex composition and the bias level show a positive coefficient sign, meaning that bearing more daughters than sons increases the probability of remaining single. All these effects are statistically significant, except for the effect of having less son percentage than ideally desired. The effect is the highest when a woman has only had daughters. Having only daughters increases the probability of remaining single by about 4.11 percentage points. However, mothers whose percentage of sons is lower than the ideal son percentage are 0.09 percentage points more likely to marry. One explanation would simply be the fact that women in this group could have at least one son, which would give them a better chance in the marriage market compared to those who have no son at all.

We also explore the effect of bearing daughters on the time that elapses between the first birth and the first union, focusing only on women who have ever been married. Hence, we estimate the equation (1.1), replacing the dependent variable by the number of years between the first birth and the first marriage or union. The results are reported in table 1-22. The findings show that having daughters when single increases marriage delay. All variables show a positive coefficient sign, the only exception being the effect of having a female firstborn but this effect is not statistically significant. In particular, having only daughters when single increases the time on the marriage market by 0.3 years, in relation to an average of 4.4 years.

Son Bias and Sexual Activity

In this last part of the mechanism, we investigate the effect of son bias on the lifetime number of sexual partners. As for the analysis of the likelihood of remaining single, we focus on women aged 45 years old or older. We estimate equation 1, where the dependent variable is the lifetime number of sexual partners, and report the results in table (1-23). We find that having more daughters or fewer sons than desired significantly increases the lifetime number of sexual partners. This is likely the result of marriage delay. The effect is consistent across different subsamples, which are the whole sample of mothers (Panel A), the subsample of mothers who have ever been married (Panel B), and the subsample mothers who have ever been married and who had their first child when they were single (Panel C). In particular, for the latter group of women, having only daughters increases the number of sexual partners, which has the largest effect. It is interesting to note that for the latter group, the average number of partners is greater than that of the group of women who experienced their first childbirth in a marital union. This reinforces the idea that for women with a high propensity to give birth to girls, those who became mothers during their celibacy spend more time in the marriage market, counting several sexual partners before getting married; which increases their likelihood of being HIV infected.

1.7 Conclusion

In this paper, we offer the first evidence of the causal effect of son bias and offspring sex composition on HIV spread. Analyzing individual level data from 37 sub-Saharan African countries, we find that son bias significantly increases the likelihood of HIV infection among women but not among men. The finding is robust across different measures of son bias and offspring sex composition.

We find that son bias is larger among patrilineal societies, where the increase in the odds to bear daughters increases the likelihood of being HIV-infected. We find that the limited ability of women in tribes where polygamy predominates to prevent their husbands from taking additional wives, as well as the physiologically more vulnerable nature of women to sexual infections, explain the stronger impact of the bias for sons on the risk of HIV infection among polygamous tribes. In tribes where bride price is mandatory, the greater impact of bearing daughters on the probability of HIV infection, in comparison to tribes where the bride price is a token or optional, allows us to conclude that the social cost associated to not bearing sons favors debauched sexual behaviour, which in turn increases the effect of bearing daughters on HIV infection risk in the first group; that is, the results suggest that value given to sons is strictly greater than the bride price. We also find evidence that richer countries display a greater level of son bias, and that the effect of son bias on HIV infection is greater in these countries.

We uncover four underlying mechanisms. First, son bias leads to larger family size, thereby increasing the likelihood of unprotected sex. Second, son bias causes wives to be unfaithful. Third, son bias causes husbands to take additional wives when the first wife is more likely to produce daughters. Fourth, women who bear a daughter when unmarried are less likely to find a husband, resulting in a larger number of sexual partners throughout their lifetime. These findings are consistent with the classical theory of sex determination according to which child sex is determined by the male sexual gametes, as well as with modern generalizations of this theory that also assign a role to mothers.

The fact that women are more vulnerable than men to HIV as a consequence of son bias is interesting, especially because son bias is much stronger among men than women in our context. Our analysis shows that a preference for sons causes women to adopt behaviors that are more to their detriment, which is a new finding. The findings suggest that policies aiming at reducing the prevalence of HIVshould integrate ways to reduce discrimination against women.

Our analysis contributes to the thin literature on the structural causes of the spread of sexually transmitted diseases. Our uncovered mechanisms are important in their own rights, and the overall findings call for policy actions in a context where HIV prevalence remains high.

1.8 Appendix

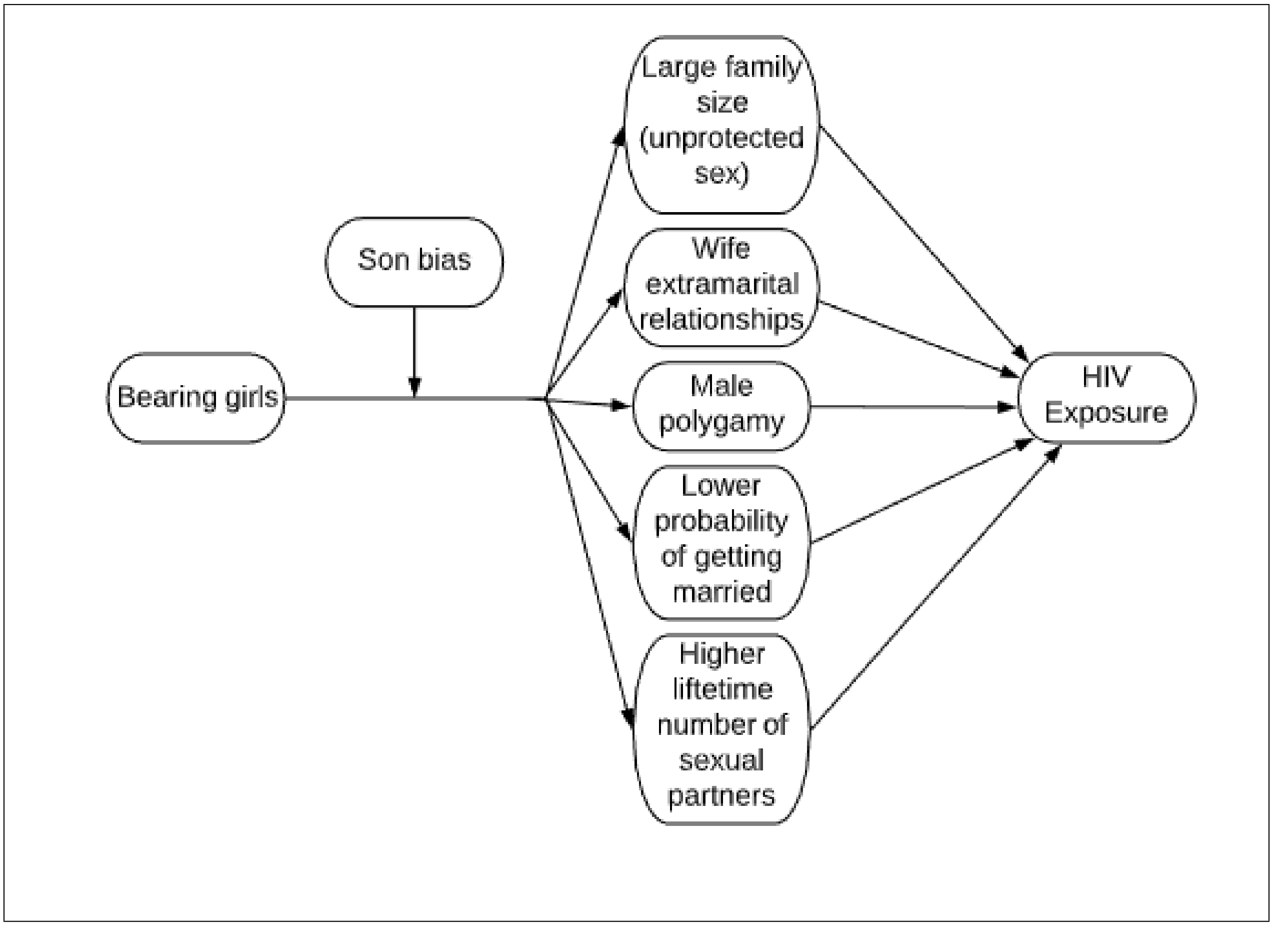
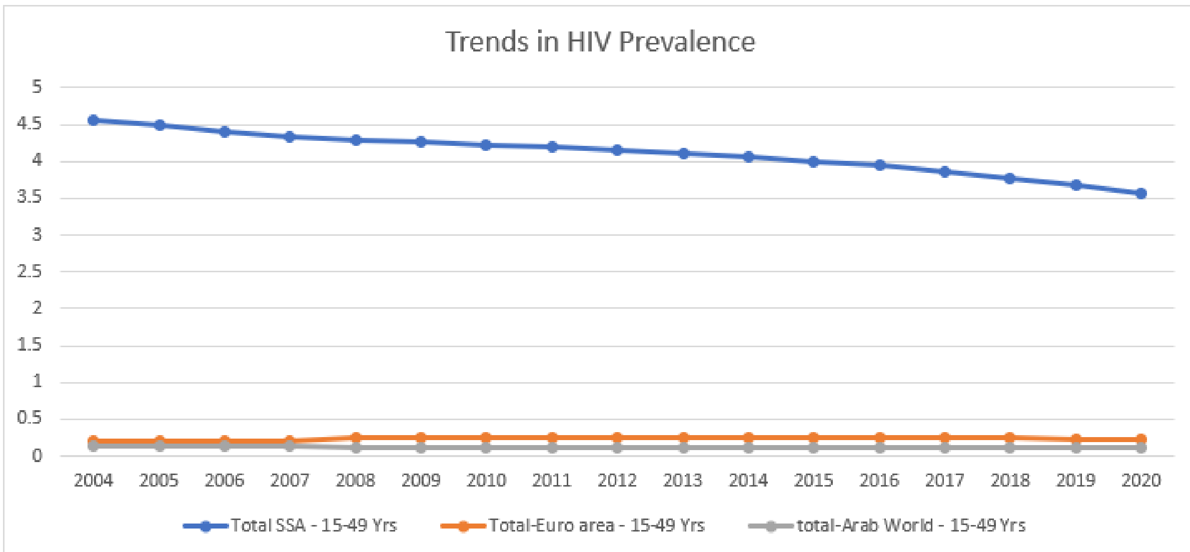


Figure 1-1: Conceptual Framework.



Source: World Bank

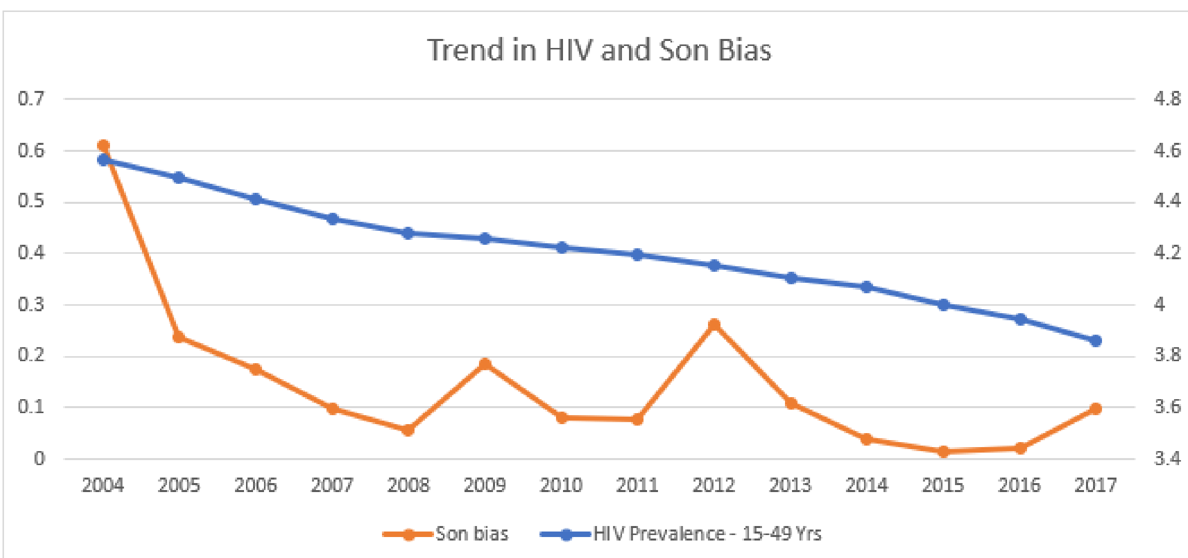


Figure 1-3: Trends in HIV and Bias among sons Biased Parents.

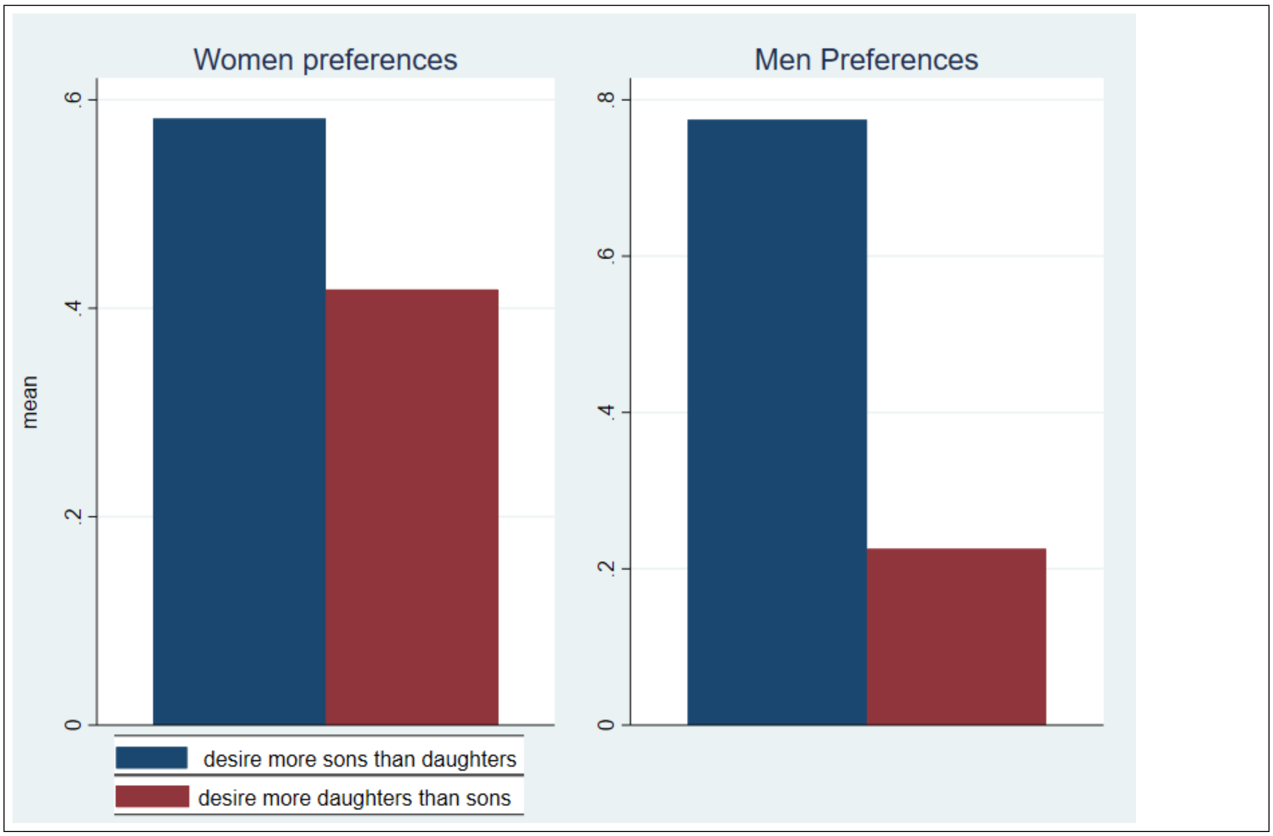


Figure 1-4: Son Bias.

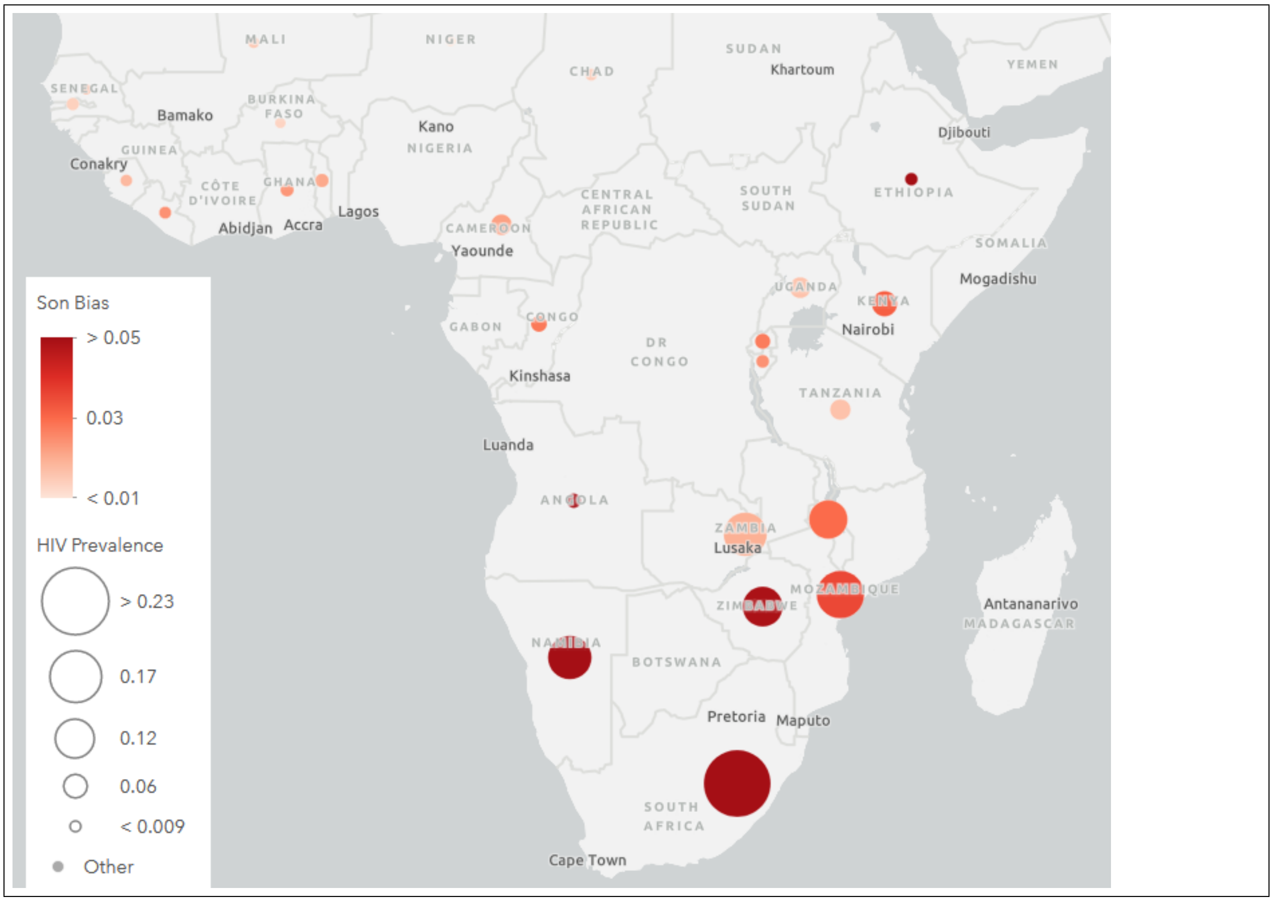


Figure 1-5: Son Bias and HIV prevalence Map.

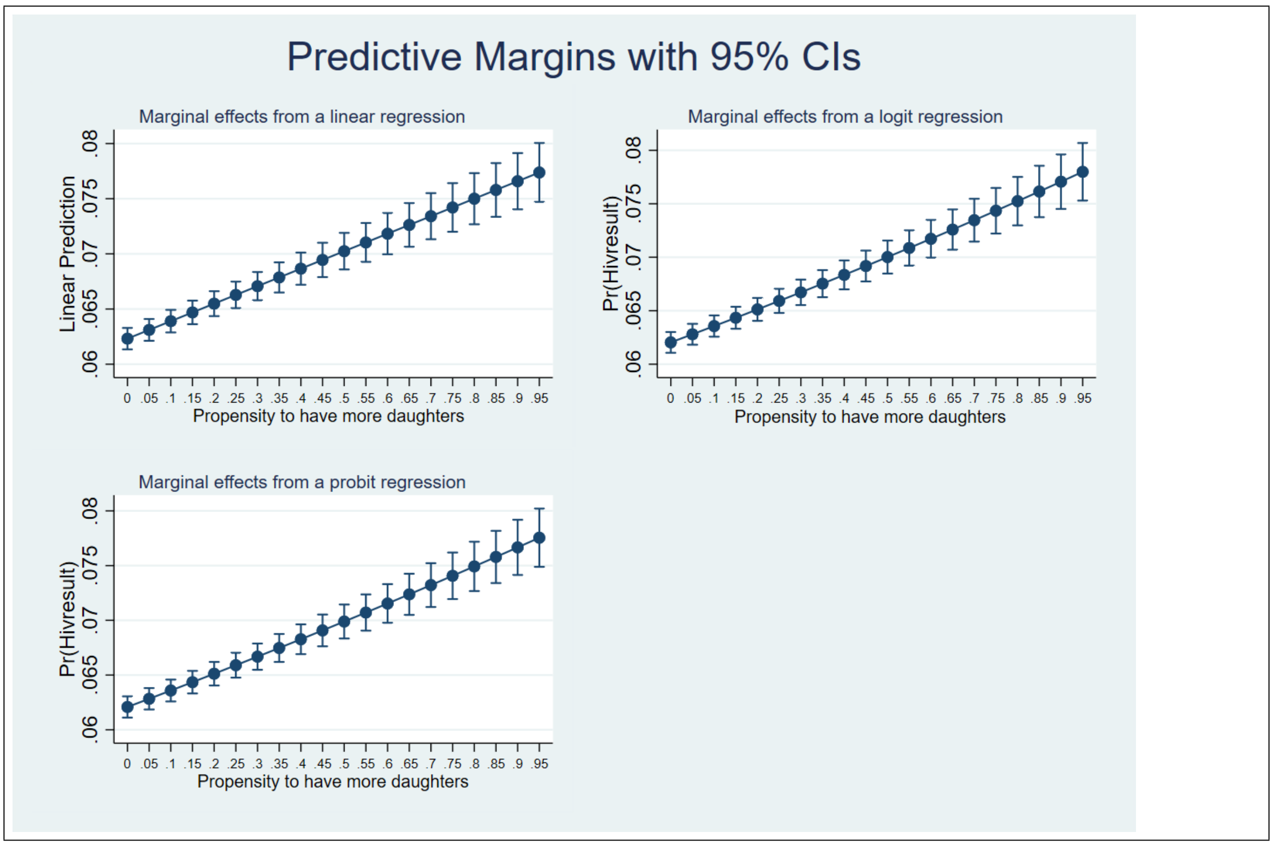


Figure 1-6: Marginal effect of son bias on HIV infection risk.

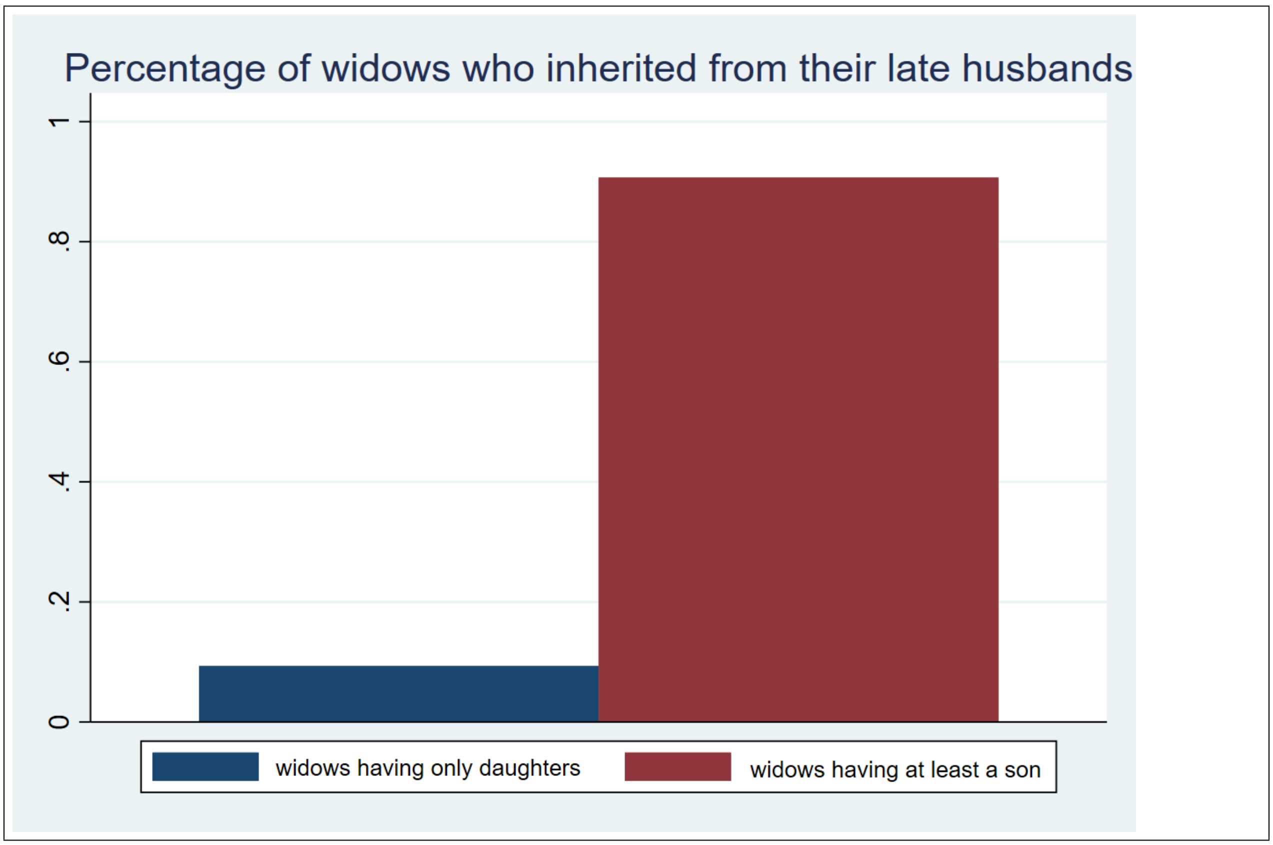


Figure 1-7: Son Bias and Widows' Inheritance.

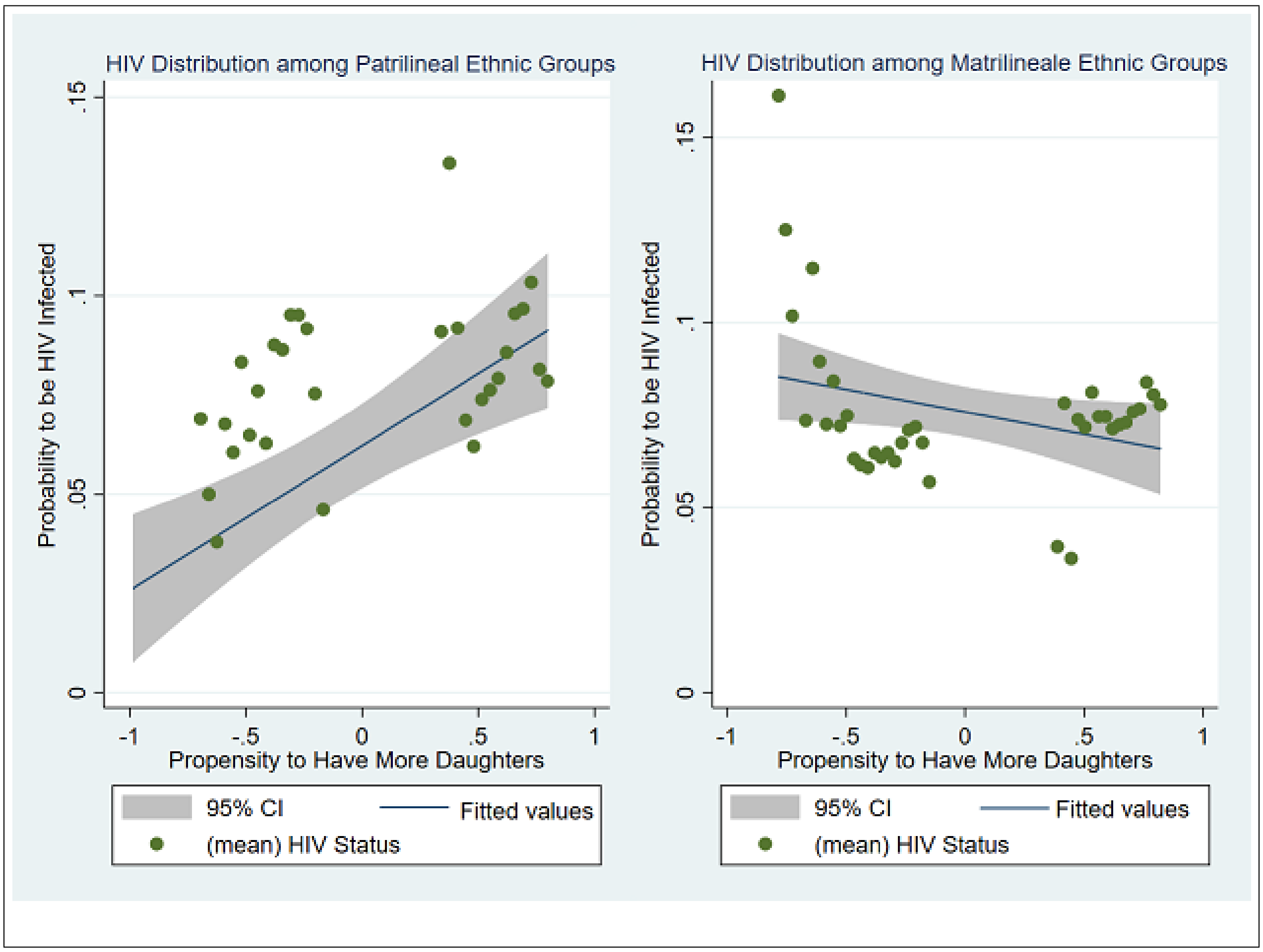


Figure 1-8: Kinship, Son Bias and HIV.

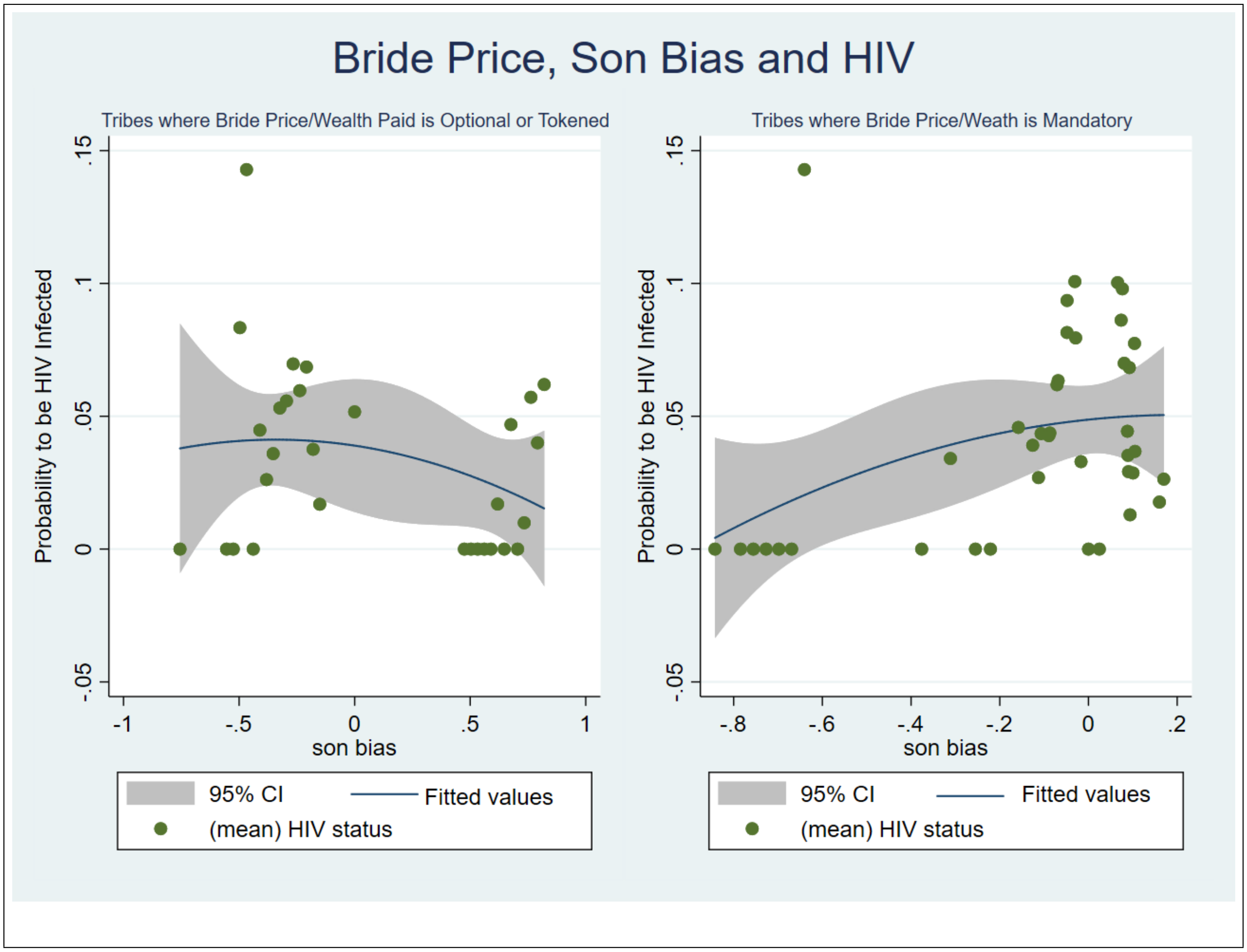
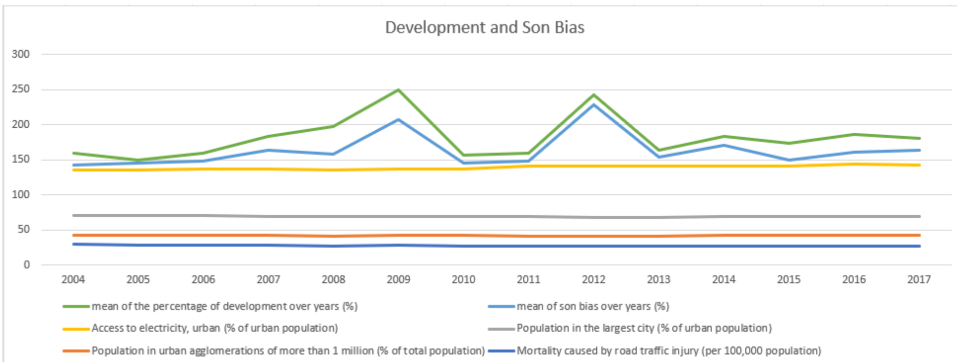


Figure 1-9: Son Bias and HIV.



Source: World Bank and Authors derived from DHS

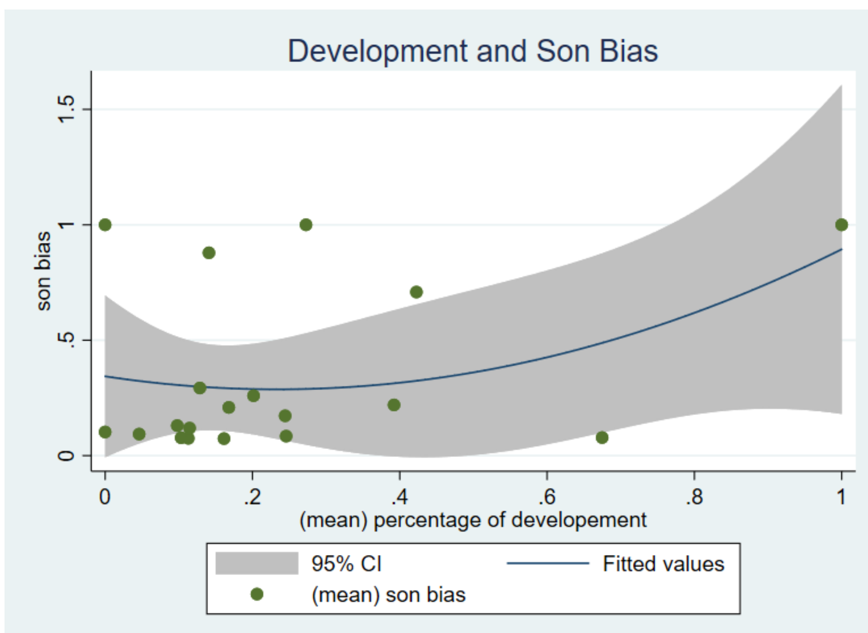


Figure 1-11: Development and Son Bias in SSA.

Table 1-1: Descriptive Statistics

	HIV surveys Respondents			***IR surveys Respondents***		
	N	Mean	Std.Dev			
HIV	264,565	0.0593	0.2382			
Mothers HIV	201,495	0.0614	0.2399			
Fathers HIV	63,070	0.0527	0.2235			
Mothers' Offspring gender composition						
First child is girl	201,495	0.4914	0.4999	368,103	0.490998	0.49992
Has only daughters	201,495	0.1681	0.374	368,103	0.190058	0.392347
Propensity to have more daughters	201,495	-0.0007	0.4878	368,103	0.093266	0.496878
Son percentage smaller than ideal son percentage	201,495	0.4656	0.4988	368,103	0.538928	0.498483
Total kids born	201,495	3.8726	2.5409	368,103	3.598126	2.488493
Mothers' characteristics						
Patrilineal tribe	201,495	0.769	0.4214	309,484	0.772757	0.419051
Matrilineal tribe	201,495	0.1851	0.3884	309,484	0.137329	0.344195
Bride price mandatory	201,495	0.8948	0.3068	368,103	0.906293	0.291422
Bride price optional or token	201,495	0.0786	0.2691	368,103	0.074754	0.262994
Polygamous tribe	201,495	0.7315	0.4432	368,103	0.078565	0.269059
Unfaithful	201,495	0.2026	0.402	368,103	0.3026	0.4594
Married at the time of survey	183,727	0.8443	0.3625	305,295	0.8258	0.3793
Never Married	183,727	0.062	0.2411	305,295	0.0659	0.2481
Ever been married	183,727	0.938	0.2411	305,295	0.9341	0.2481
Polygamous	201,495	0.4413	0.4965	368,103	0.5276	0.4992
Primary education	201,495	0.3435	0.4749	368,103	0.3407	0.474
higher education	201,495	0.2289	0.4201	368,103	0.2628	0.4402
age	201,495	31.7724	8.4447	368,103	31.473	8.4192
Lifetime sexual partners	145,437	3.3191	11.2715	242,010	3.1907	10.8419
Survey years	201,495			368,103		
Fathers' Offspring gender composition						
First child is girl	63,070	0.4914	0.4999			
Has only daughters	63,070	0.1445	0.3516			
Propensity to have more daughters	63,070	0.0004	0.4832			
Son percentage smaller than ideal son percentage	63,070	0.4236	0.4885			
Total kids born	63,070	4.1582	2.5286			
Fathers' characteristics						
Patrilineal tribe	62,124	0.7532	0.4298			
Matrilineal tribe	62,124	0.1959	0.3969			
Bride price mandatory	63,070	0.8806	0.3243			
Bride price optional or token	63,070	0.0868	0.2815			
Polygamous tribe	63,070	0.7042	0.4564			
Unfaithful	63,070	0.1417	0.3487			
age	63,070	31.4281	8.5263			
Primary education	63,070	0.3428	0.4747			
higher education	63,070	0.2435	0.4292			
Survey years	63,070					

Table 1-2: Son Bias and HIV

Panel A: Sample of mothers

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	0.00332***			
	(0.000983)			
propensity to have more daughters		0.00445***		
		(0.00124)		
has only daughters			0.00600***	
			(0.00138)	
son percentage smaller than ideal son percentage				0.00212*
				(0.00125)
N	201495	201495	201495	201495
R-sq	0.0599	0.0600	0.0600	0.0599
ymean	0.0615	0.0615	0.0615	0.0615

Panel B: Sample of fathers

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	-0.00671***			
	(0.00224)			
propensity to have more daughters		-0.00282**		
		(0.00126)		
has only daughters			-0.00178	
			(0.00267)	
son percentage smaller than ideal son percentage				-0.00158
				(0.00173)
N	63070	63070	63070	63070
R-sq	0.107	0.107	0.107	0.107
ymean	0.0528	0.0528	0.0528	0.0528

Panel C: Mothers married to fathers in Panel B

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	0.00383**			
	(0.00162)			
propensity to have more daughters		0.00442**		
		(0.00199)		
has only daughters			0.00828**	
			(0.00335)	
son percentage smaller than ideal son percentage				0.000854
				(0.00176)
N	60648	60648	60648	60648
R-sq	0.102	0.102	0.102	0.102
ymean	0.0543	0.0543	0.0543	0.0543
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnicity by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, $*** * * p < 0.05, *p < 0.1$. The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-3: Son Bias and HIV: Heterogeneity by Education Level (***Mothers***)

Panel A: Mothers did not go further than primary

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	0.00290**			
	(0.00113)			
propensity to have more daughters		0.00298**		
		(0.00133)		
has only daughters			0.00560***	
			(0.00148)	
son percentage smaller than ideal son percentage				0.00119
				(0.00143)
N	155368	155368	155368	155368
R-sq	0.0561	0.0561	0.0561	0.0560
ymean	0.0545	0.0545	0.0545	0.0545

Panel B: Mothers did at least secondary school

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	0.00470			
	(0.00335)			
propensity to have more daughters		0.00904**		
		(0.00343)		
has only daughters			0.00719**	
			(0.00345)	
son percentage smaller than ideal son percentage				0.00556*
				(0.00322)
N	46127	46127	46127	46127
R-sq	0.0973	0.0974	0.0973	0.0973
ymean	0.0841	0.0841	0.0841	0.0841
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✗	✗	✗	✗
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnicity by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $*** \ p < 0.01$, $*** * * p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-4: Son Bias and HIV: Heterogeneity by Education Level (***Fathers***)

Panel A: Fathers did not go further than primary

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	-0.00486**			
	(0.00200)			
propensity to have more daughters		0.000573		
		(0.00144)		
has only daughters			0.00473*	
			(0.00260)	
son percentage smaller than ideal son percentage				0.00192
				(0.00169)
N	41989	41989	41989	41989
R-sq	0.111	0.111	0.111	0.111
ymean	0.0415	0.0415	0.0415	0.0415

Panel B: Fathers did at least secondary school

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	-0.0102**			
	(0.00500)			
propensity to have more daughters		-0.0184***		
		(0.00680)		
has only daughters			-0.0135	
			(0.00977)	
son percentage smaller than ideal son percentage				-0.0140***
				(0.00525)
N	21079	21079	21079	21079
R-sq	0.138	0.138	0.138	0.138
ymean	0.0754	0.0754	0.0754	0.0754
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✗	✗	✗	✗
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnicity by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $* * p < 0.05, * p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-5: Son Bias and HIV: Heterogeneity by Age at First Birth (***Mothers***)

First birth between 15 to 24 years old

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	0.00230**			
	(0.00107)			
propensity to have more daughters		0.00411***		
		(0.00137)		
has only daughters			0.00486***	
			(0.00176)	
son percentage smaller than ideal son percentage				0.00173
				(0.00139)
N	161184	161184	161184	161184
R-sq	0.0642	0.0642	0.0642	0.0642
ymean	0.0625	0.0625	0.0625	0.0625

First birth between 25 to 39 years old

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	0.00863*			
	(0.00442)			
propensity to have more daughters		0.0101**		
		(0.00499)		
has only daughters			0.0198***	
			(0.00506)	
son percentage smaller than ideal son percentage				0.00764
				(0.00513)
N	17469	17469	17469	17469
R-sq	0.154	0.154	0.155	0.154
ymean	0.0551	0.0551	0.0551	0.0551

First birth at 40 years or older

Variables	(1)	(2)	(3)	(4)
	HIV (0/1)			
first child is girl	0.00866*			
	(0.00488)			
propensity to have more daughters		0.00851		
		(0.00551)		
has only daughters			0.0129**	
			(0.00530)	
son percentage smaller than ideal son percentage				0.191
				(0.128)
N	9850	9850	9850	9850
R-sq	0.112	0.112	0.113	0.112
ymean	0.0515	0.0515	0.0515	0.0515
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnicity by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-6: Gender Composition and HIV: Heterogeneity by Gender Preference

Panel A: Mothers with bias towards daughters

	(1)	(2)	(3)	(4)
		HIV (0/1)		
Variables				
first child is girl	0.00300			
	(0.00379)			
propensity to have more daughters		0.00464		
		(0.0308)		
has only daughters			0.0144**	
			(0.00612)	
son percentage smaller than ideal son percentage				-0.00440
				(0.00500)
N	21136	21136	21136	21136
R-sq	0.121	0.121	0.121	0.121
ymean	0.0492	0.0492	0.0492	0.0492

Panel B: Mothers with bias towards sons

	(1)	(2)	(3)	(4)
		HIV (0/1)		
Variables				
first child is girl	0.00412***			
	(0.00149)			
propensity to have more daughters		0.00625**		
		(0.00303)		
has only daughters			0.0114**	
			(0.00546)	
son percentage smaller than ideal son percentage				0.00404*
				(0.00239)
N	119649	119649	119649	119649
R-sq	0.0670	0.0670	0.0670	0.0670
ymean	0.0642	0.0642	0.0642	0.0642
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnicity by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-7: Son Bias and HIV: Robustness Check

	(1)	(2)	(3)	(4)
	HIV (0/1)			
Variables	**Sample of mothers with female firstborn**	**Sample of mothers with positive propensity to bear daughters**	**Sample of mothers having only daughters**	**Sample of mothers who have more sons than desired**
Son bias	0.000713***	0.000753***	0.000678***	0.000559**
	(0.000222)	(0.000256)	(0.000249)	(0.000225)
N	92264	74580	91955	86166
R-sq	0.0735	0.0780	0.0725	0.0716
ymean	0.0643	0.0646	0.0615	0.0588
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not. The variable son bias is defined as suggested by Jayachandran (2017b). It represents the residuals of regressing individuals' ideal number of desired sons on their ideal total number of desired children.*

Table 1-8: Patrilineal/Matrilineal kinship, Son Bias and HIV

Panel A: Mothers in patrilineal tribes

	(1)	(2)	(3)	(4)
	HIV (0/1)			
Variables				
first child is girl	0.00304**			
	(0.00117)			
propensity to have more daughters		0.00456***		
		(0.00161)		
has only daughters			0.00661***	
			(0.00155)	
son percentage smaller than ideal son percentage				0.000726***
				(0.000211)
N	154959	154959	154959	154959
R-sq	0.0539	0.0539	0.0540	0.0538
ymean	0.0527	0.0527	0.0527	0.0527

Panel B: Mothers in matrilineal tribes

	(1)	(2)	(3)	(4)
	HIV (0/1)			
Variables				
first child is girl	0.00262			
	(0.00257)			
propensity to have more daughters		-0.00685***		
		(0.00213)		
has only daughters			0.00242	
			(0.00405)	
son percentage smaller than ideal son percentage				0.000354
				(0.000383)
N	37295	37295	37295	37295
R-sq	0.0794	0.0795	0.0794	0.0794
ymean	0.0784	0.0784	0.0784	0.0784
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, $*** * * p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not*

Table 1-9: Polygamous Tribes, Son Bias and HIV

Panel A: Mothers in tribes where polygamy predominates

Variables	(1)	(2)	(3) HIV (0/1)	(4)
first child is girl	0.00403***			
	(0.00133)			
propensity to have more daughters		0.00645***		
		(0.00154)		
has only daughters			0.00749***	
			(0.00154)	
son percentage smaller than ideal son percentage				0.00367***
				(0.00137)
N	147388	147388	147388	147388
R-sq	0.0556	0.0557	0.0557	0.0556
ymean	0.0597	0.0597	0.0597	0.0597

Panel B: Mothers in tribes where monogamy predominates

Variables	(1)	(2)	(3) HIV (0/1)	(4)
first child is girl	0.00205			
	(0.00232)			
propensity to have more daughters		0.000179		
		(0.00212)		
has only daughters			0.00492*	
			(0.00274)	
son percentage smaller than ideal son percentage				-0.00334
				(0.00309)
N	40589	40589	40589	40589
R-sq	0.102	0.102	0.102	0.102
ymean	0.0558	0.0558	0.0558	0.0558
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $*** \, p < 0.01$, $*** * * p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-10: Marriage Mode, Son Bias and HIV

	Panel A: Mothers in tribes with optional bride price			
	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	-0.00649			
	(0.00756)			
propensity to have more daughters		-0.00113		
		(0.00402)		
has only daughters			0.00334	
			(0.00297)	
son percentage smaller than ideal son percentage				-0.00659
				(0.00446)
N	15837	15837	15837	15837
R-sq	0.0894	0.0892	0.0893	0.0894
ymean	0.0544	0.0544	0.0544	0.0544

	Panel B: Mothers in tribes with mandatory bride price			
	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	0.00416***			
	(0.00103)			
propensity to have more daughters		0.00516***		
		(0.00133)		
has only daughters			0.00638***	
			(0.00158)	
son percentage smaller than ideal son percentage				0.00261**
				(0.00121)
N	180294	180294	180294	180294
R-sq	0.0572	0.0572	0.0572	0.0571
ymean	0.0594	0.0594	0.0594	0.0594
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-11: Marriage Mode, kinship, Son Bias and HIV

Panel A: Mothers in patrilineal tribes with mandatory bride price

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	0.00376***			
	(0.00125)			
propensity to have more daughters		0.00494***		
		(0.00172)		
has only daughters			0.00684***	
			(0.00175)	
son percentage smaller than ideal son percentage				0.00227
				(0.00152)
N	140277	140277	140277	140277
R-sq	0.0563	0.0564	0.0564	0.0563
ymean	0.0535	0.0535	0.0535	0.0535

Panel B: Mothers in matrilineal tribes with mandatory bride price

	(1)	(2)	(3)	(4)
			HIV (0/1)	
Variables				
first child is girl	0.00270			
	(0.00429)			
propensity to have more daughters		0.00311		
		(0.00400)		
has only daughters			-0.0027	
			(0.00429)	
son percentage smaller than ideal son percentage				-0.00514
				(0.00701)
N	33896	33896	33896	33896
R-sq	0.0807	0.0808	0.0806	0.0807
ymean	0.0794	0.0794	0.0794	0.0794
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $^{***}p < 0.01$, $^{***} ** p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-12: Development, Son Bias and HIV

Panel A: Mothers in more developed countries

	(1)	(2)	(3)	(4)
		HIV (0/1)		
Variables				
first child is girl	0.00295			
	(0.00308)			
propensity to have more daughters		0.00911**		
		(0.00384)		
has only daughters			0.0149***	
			(0.00387)	
son percentage smaller than ideal son percentage				0.00113
				(0.00319)
N	46459	46459	46459	46459
R-sq	0.0624	0.0626	0.0628	0.0624
ymean	0.103	0.103	0.103	0.103

Panel B: Mothers in less developed countries

	(1)	(2)	(3)	(4)
		HIV (0/1)		
Variables				
first child is girl	0.00329***			
	(0.00106)			
propensity to have more daughters		0.00295**		
		(0.00132)		
has only daughters			0.00321**	
			(0.00134)	
son percentage smaller than ideal son percentage				0.0101**
				(0.00312)
N	155036	155036	155036	155036
R-sq	0.0476	0.0475	0.0475	0.0475
ymean	0.0492	0.0492	0.0492	0.0492
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. The dependent variable takes the value one if the individual is infected and 0 if not.*

Table 1-13: Mechanism: Son Bias and Family Size

	Participants to HIV surveys				Participants to IR surveys			
Panel A: All mothers								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Total children ever born				Total children ever born			
Variables	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy
first child is girl	0.0256**	0.0317**	0.0283	0.0684**	0.0272***	0.0324***	0.0302*	0.0495**
	(0.0119)	(0.0125)	(0.0242)	(0.0310)	(0.00656)	(0.00836)	(0.0173)	(0.0247)
N	201495	172338	11389	22526	368103	285180	20115	37311
R-sq	0.508	0.508	0.460	0.467	0.496	0.506	0.406	0.475
ymean	3.845	4.088	1.677	4.101	3.550	3.947	1.572	4.003
Panel B: Mothers with bias for sons								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Total children ever born				Total children ever born			
Variables	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy
first child is girl	0.0352**	0.0395**	0.0191	0.116**	0.0187**	0.0206	0.0416**	0.0472
	(0.0140)	(0.0149)	(0.0302)	(0.0446)	(0.00875)	(0.0129)	(0.0202)	(0.0354)
N	119649	100832	7387	13101	241644	166391	12446	22001
R-sq	0.504	0.505	0.456	0.503	0.478	0.490	0.391	0.471
ymean	3.778	4.042	1.634	4.023	3.340	3.839	1.537	3.860
Panel C: Mothers with bias for daughters								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Total children ever born				Total children ever born			
Variables	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy	All mothers	Ever been married mothers	Never maried mothers	Mothers who first birth was during celibacy
first child is girl	-0.00274	0.00386	-0.00852	-0.0259	0.0154	0.0186*	-0.0171	0.0329
	(0.0155)	(0.0166)	(0.0415)	(0.0361)	(0.0105)	(0.0103)	(0.0249)	(0.0342)
N	81846	71506	4002	9425	126459	118789	7669	15310
R-sq	0.532	0.531	0.564	0.506	0.570	0.556	0.512	0.543
ymean	3.942	4.155	1.756	4.212	3.955	4.101	1.632	4.216
Age	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓
Wealth	✓	✓	✓	✓	✓	✓	✓	✓
Religion	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable is a continuous variable that captures the number of children ever born by the individual.*

Table 1-14: Mechanism: Son Bias and Desire for an Additional Child

	Participants to HIV surveys				**Participants to IR surveys**			
	Sample of mothers with more children than desired							
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Desire more children (0/1)				**Desire more children (0/1)**			
Variables								
first child is girl	0.00497				0.00693*			
	(0.00403)				(0.00386)			
propensity to have more daughters		0.0227***				0.0123***		
		(0.00327)				(0.00257)		
has only daughters			0.0565***				0.0442***	
			(0.00491)				(0.00830)	
son percentage smaller than ideal son percentage				0.0329***				0.0225***
				(0.00375)				(0.00381)
N	59942	59942	59942	59942	88153	88153	88153	88153
R-sq	0.292	0.292	0.293	0.293	0.154	0.155	0.155	0.155
ymean	0.338	0.338	0.338	0.338	0.192	0.192	0.192	0.192
Age	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓
Wealth	✓	✓	✓	✓	✓	✓	✓	✓
Religion	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the individual is willing to have other children, and the value zero if not.*

Table 1-15: Mechanism: Son Bias and Wife Unfaithfulness

	All currently married mothers				Mothers whose husbands have bias toward sons			
Panel A: All currently married mothers								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	Unfaithful during last 12 months (0/1)				Unfaithful during last 12 months (0/1)			
first child is girl	-0.000761				-0.000714			
	(0.00153)				(0.00361)			
propensity to have more daughters		0.00127				0.00989**		
		(0.00137)				(0.00378)		
has only daughters			0.00182				0.0238***	
			(0.00176)				(0.00517)	
son percentage smaller than ideal son percentage				0.0589***				0.0372***
				(0.0165)				(0.00814)
N	252114	252114	252114	252114	32586	32586	32586	32586
R-sq	0.281	0.281	0.281	0.291	0.527	0.527	0.527	0.527
ymean	0.0912	0.0912	0.0912	0.0912	0.103	0.103	0.103	0.103
Panel B: Mothers married for less than 5 years								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	Unfaithful during last 12 months (0/1)				Unfaithful during last 12 months (0/1)			
first child is girl	-0.00386				0.000391			
	(0.00286)				(0.00801)			
propensity to have more daughters		-0.00419				0.00481		
		(0.00307)				(0.00811)		
has only daughters			-0.00381				0.00491	
			(0.00311)				(0.00950)	
son percentage smaller than ideal son percentage				0.0598***				0.0272*
				(0.0189)				(0.0149)
N	49126	49126	49126	49126	5831	5831	5831	5831
R-sq	0.266	0.266	0.266	0.274	0.625	0.625	0.625	0.625
ymean	0.117	0.117	0.117	0.117	0.0726	0.0726	0.0726	0.0726
Panel C: Mothers married for 5 to 9 years								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	Unfaithful during last 12 months (0/1)				Unfaithful during last 12 months (0/1)			
first child is girl	0.000158				-0.00227			
	(0.00307)				(0.00738)			
propensity to have more daughters		0.00583**				0.0136*		
		(0.00247)				(0.00801)		
has only daughters			0.0166***				0.00871	
			(0.00275)				(0.00906)	
son percentage smaller than ideal son percentage				0.0577***				0.0235
				(0.0189)				(0.0188)
N	61027	61027	61027	61027	7368	7368	7368	7368
R-sq	0.251	0.251	0.251	0.258	0.651	0.651	0.651	0.651
ymean	0.119	0.119	0.119	0.119	0.0835	0.0835	0.0835	0.0835
Age	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓
Wealth	✓	✓	✓	✓	✓	✓	✓	✓
Religion	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. The dependent variable capture mothers' unfaithfulness.*

Table 1-16: Mechanism: Son Bias and Wife Unfaithfulness (2)

	All currently married mothers				Mothers whose husbands have bias toward sons			
	Panel D: Mothers married for 10 to 14 years							
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	Unfaithful during last 12 months (0/1)				Unfaithful during last 12 months (0/1)			
first child is girl	0.00458 (0.00328)				0.00646 (0.00563)			
propensity to have more daughters		0.00324 (0.00343)				0.00259 (0.00914)		
has only daughters			0.0193*** (0.00628)				0.00159 (0.0155)	
son percentage smaller than ideal son percentage				0.0594*** (0.0185)				0.00601 (0.0152)
N	24230	24230	24230	24230	6395	6395	6395	6395
R-sq	0.259	0.259	0.259	0.267	0.650	0.650	0.650	0.650
ymean	0.117	0.117	0.117	0.117	0.0655	0.0655	0.0655	0.0655
	Panel E: Mothers married for at least 15 years							
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Variables	Unfaithful during last 12 months (0/1)				Unfaithful during last 12 months (0/1)			
first child is girl	0.00458 (0.00197)				0.000885 (0.00535)			
propensity to have more daughters		0.00580*** (0.00165)				0.00877* (0.00509)		
has only daughters			0.0207*** (0.00541)				0.0231* (0.0135)	
son percentage smaller than ideal son percentage				0.0482*** (0.0135)				0.0386** (0.0159)
N	117731	117731	117731	117731	11519	11519	11519	11519
R-sq	0.231	0.231	0.231	0.236	0.505	0.505	0.505	0.505
ymean	0.110	0.110	0.110	0.110	0.0600	0.0600	0.0600	0.0600
Age	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓
Wealth	✓	✓	✓	✓	✓	✓	✓	✓
Religion	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $*** \; p < 0.01$, $*** ** p < 0.05, *p < 0.1$. *The dependent variable capture mothers' unfaithfulness.*

Table 1-17: Mechanism: Son Bias and Husband Unfaithfulness

Panel A: All currently married fathers

	(1)	(2)	(3)	(4)
	Unfaithful during last 12 months (0/1)			
Variables				
first child is girl	-0.00377			
	(0.00270)			
propensity to have more daughters		-0.00208		
		(0.00335)		
has only daughters			-0.00928**	
			(0.00438)	
son percentage smaller than ideal son percentage				0.000340
				(0.00323)
N	63885	63885	63885	63885
R-sq	0.114	0.114	0.114	0.114
ymean	0.108	0.108	0.108	0.108

Panel B: Fathers married for less than 5 years

	(1)	(2)	(3)	(4)
	Unfaithful during last 12 months (0/1)			
Variables				
first child is girl	0.000278			
	(0.00793)			
propensity to have more daughters		-0.000463		
		(0.00863)		
has only daughters			-0.00614	
			(0.00840)	
son percentage smaller than ideal son percentage				0.00154
				(0.00912)
N	9994	9994	9994	9994
R-sq	0.178	0.178	0.178	0.178
ymean	0.126	0.126	0.126	0.126

Panel C: Fathers married for 5 to 9 years

	(1)	(2)	(3)	(4)
	Unfaithful during last 12 months (0/1)			
Variables				
first child is girl	-0.00883			
	(0.00898)			
propensity to have more daughters		-0.0129		
		(0.00788)		
has only daughters			-0.0152**	
			(0.00755)	
son percentage smaller than ideal son percentage				-0.00630
				(0.0752)
N	12087	12087	12087	12087
R-sq	0.171	0.171	0.171	0.171
ymean	0.128	0.128	0.128	0.128
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the ethnic group level. **** $p < 0.01$, ***** $**p < 0.05, *p < 0.1$. *The dependent variable capture fathers' unfaithfulness.*

Table 1-18: Mechanism: Son Bias and Husband Unfaithfulness (2)

Panel D: Fathers married for 10 to 14 years

	(1)	(2)	(3)	(4)
	Unfaithful during last 12 months (0/1)			
Variables				
first child is girl	-0.00409			
	(0.00703)			
propensity to have more daughters		0.00421		
		(0.00675)		
has only daughters			-0.00107	
			(0.00986)	
son percentage smaller than ideal son percentage				-0.00734
				(0.00668)
N	11112	11112	11112	11112
R-sq	0.178	0.178	0.178	0.179
ymean	0.115	0.115	0.115	0.115

Panel E: Fathers married for at least 15 years

	(1)	(2)	(3)	(4)
	Unfaithful during last 12 months (0/1)			
Variables				
first child is girl	-0.00199			
	(0.00374)			
propensity to have more daughters		-0.00186		
		(0.00576)		
has only daughters			-0.00826	
			(0.00865)	
son percentage smaller than ideal son percentage				0.00150
				(0.00452)
N	26040	26040	26040	26040
R-sq	0.127	0.127	0.127	0.127
ymean	0.0919	0.0919	0.0919	0.0919
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable capture fathers' unfaithfulness.*

Table 1-19: Mechanism: Son Bias and Polygamy (1)

	All currently married mothers				All mothers whose husbands have bias toward sons			
Panel A: All currently married mothers								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Polygamy (0/1)				Polygamy (0/1)			
Variables								
first child is girl	0.000130				0.0182***			
	(0.00195)				(0.00668)			
propensity to have more daughters		0.0105***				0.0260***		
		(0.00355)				(0.00731)		
has only daughters			0.0214***				0.0328***	
			(0.00471)				(0.00682)	
son percentage smaller than ideal son percentage				0.0524***				0.0372***
				(0.0120)				(0.00814)
N	285180	285180	285180	285180	32586	32586	32586	32586
R-sq	0.124	0.124	0.125	0.127	0.202	0.202	0.202	0.203
ymean	0.387	0.387	0.387	0.387	0.269	0.269	0.269	0.269
Panel B: Married mothers aged 15 to 29 years old								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Polygamy (0/1)				Polygamy (0/1)			
Variables								
first child is girl	0.00175				0.0219**			
	(0.00331)				(0.0107)			
propensity to have more daughters		0.00641*				0.00778		
		(0.00364)				(0.00828)		
has only daughters			0.00739				0.00968	
			(0.00481)				(0.0105)	
son percentage smaller than ideal son percentage				0.0564***				0.0183*
				(0.0143)				(0.00931)
N	117271	117271	117271	117271	15300	15300	15300	15300
R-sq	0.118	0.118	0.118	0.122	0.236	0.236	0.236	0.236
ymean	0.305	0.305	0.305	0.305	0.247	0.247	0.247	0.247
Panel C: Married mothers aged 30 to 44 years old								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Polygamy (0/1)				Polygamy (0/1)			
Variables								
first child is girl	0.00140				0.00900			
	(0.00338)				(0.00823)			
propensity to have more daughters		0.0128**				0.0343***		
		(0.00493)				(0.0112)		
has only daughters			0.0461***				0.0640***	
			(0.00640)				(0.0222)	
son percentage smaller than ideal son percentage				0.0503***				0.0465***
				(0.0121)				(0.0128)
N	137507	137507	137507	137507	14987	14987	14987	14987
R-sq	0.116	0.116	0.116	0.118	0.265	0.266	0.266	0.267
ymean	0.426	0.426	0.426	0.426	0.281	0.281	0.281	0.281
Panel D: Married mothers aged 45 or more								
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Polygamy (0/1)				Polygamy (0/1)			
Variables								
first child is girl	0.00428				0.0422			
	(0.00521)				(0.0422)			
propensity to have more daughters		0.0231***				0.0507		
		(0.00709)				(0.0307)		
has only daughters			0.0904***				0.162***	
			(0.0188)				(0.0536)	
son percentage smaller than ideal son percentage				0.0448***				0.0804***
				(0.0109)				(0.0301)
N	30402	30402	30402	30402	2299	2299	2299	2299
R-sq	0.137	0.137	0.138	0.138	0.449	0.449	0.451	0.452
ymean	0.527	0.527	0.527	0.527	0.339	0.339	0.339	0.339
Age	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓
Wealth	✓	✓	✓	✓	✓	✓	✓	✓
Religion	✓	✓	✓	✓	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. The dependent variable indicates if the mother is in a polygamous union.*

Table 1-20: Mechanism: Son Bias and Polygamy (First Wives)

Mothers aged 45 plus whose husbands have bias toward sons

	(1)	(2)	(3)	(4)
		Polygamy (0/1)		
Variables				
first child is girl	0.0321*			
	(0.0169)			
propensity to have more daughters		0.00797		
		(0.0196)		
has only daughters			0.132***	
			(0.0471)	
son percentage smaller than ideal son percentage				0.0144**
				(0.00620)
N	4025	4025	4099	4025
R-sq	0.339	0.338	0.340	0.341
ymean	0.199	0.199	0.198	0.199
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $** p < 0.05, *p < 0.1$. *The dependent variable indicates if the mother is in a polygamous union.*

Table 1-21: Mechanism: Son Bias and Likelihood of Never Getting Married

	Sample of mothers aged 45 years old or more			
	(1)	(2)	(3)	(4)
	Never Married (0/1)			
Variables				
first child is girl	0.00075			
	(0.00198)			
propensity to have more daughters		0.0049**		
		(0.0019)		
has only daughters			0.0411***	
			(0.00665)	
son percentage smaller than ideal son percentage				-0.0009***
				(0.00174)
N	30898	30898	30898	30898
R-sq	0.082	0.082	0.087	0.082
ymean	0.016	0.016	0.016	0.016
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $**p < 0.05, *p < 0.1$. *The dependent variable takes the value one if the mother has never been married and the value zero if she have ever been married.*

Table 1-22: Mechanism: Son Bias and Duration in the Marriage Market

	Ever married mothers with first birth during celibacy			
	(1)	(2)	(3)	(4)
	Years from first birth to first marriage			
Variables				
first child is girl	-0.0629			
	(0.0433)			
propensity to have more daughters		0.122**		
		(0.0538)		
has only daughters			0.289***	
			(0.0649)	
son percentage smaller than ideal son percentage				0.0193***
				(0.00481)
N	37311	37311	37311	37311
R-sq	0.164	0.164	0.164	0.164
ymean	4.405	4.405	4.405	4.405
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. *** $p < 0.01$, *** $** p < 0.05, *p < 0.1$. *The dependent variable is the number of years between the first firth and the first marriage.*

Table 1-23: Mechanism: Son Bias and Sexual Behaviour

Panel A: All mothers sample

Variables	(1)	(2)	(3)	(4)
	Lifetime number of sexual partners			
first child is girl	-0.00601			
	(0.0109)			
propensity to have more daughters		0.0216*		
		(0.0115)		
has only daughters			0.0221**	
			(0.00851)	
son percentage smaller than ideal son percentage				0.0160
				(0.0517)
N	242010	242010	242010	242010
R-sq	0.0627	0.0626	0.0626	0.0626
ymean	3.249	3.249	3.249	3.249

Panel B: All married mothers

Variables	(1)	(2)	(3)	(4)
	Lifetime number of sexual partners			
first child is girl	-0.00691			
	(0.00677)			
propensity to have more daughters		0.0236***		
		(0.00675)		
has only daughters			0.0472***	
			(0.0102)	
son percentage smaller than ideal son percentage				0.0209***
				(0.00762)
N	222145	222145	222145	222145
R-sq	0.210	0.210	0.210	0.210
ymean	1.895	1.895	1.895	1.895

Panel C: Mothers aged 45+ with birth during celibacy

Variables	(1)	(2)	(3)	(4)
	Lifetime number of sexual partners			
first child is girl	0.00132			
	(0.0225)			
propensity to have more daughters		0.101***		
		(0.0277)		
has only daughters			0.188***	
			(0.0550)	
son percentage smaller than ideal son percentage				0.0654***
				(0.0222)
N	23801	23801	23801	23801
R-sq	0.266	0.267	0.266	0.266
ymean	2.030	2.030	2.030	2.030
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $*** \ p < 0.01$, $*** * * p < 0.05, *p < 0.1$. *The dependent variable is a continuous variable representing the lifetime number of sexual partners of mothers.*

Table A.1: List of surveys

Countries	DHS Years	HIV Years
Angola	2006/2007, 2011, 2015/2016	2006, 2011, 2015
Burkina Faso	1992,1993, 1998, 1999, 2003, 2010, 2014	2003, 2010, 2014
Benin	1996, 2001, 2006, 2011, 2012	2001, 2012
Burundi	1987, 2010, 2011, 2012, 2013, 2016, 2017	2010, 2012, 2016,
Congo	2007, 2013, 2014	2007, 2013, 2014
Central Africa Republic	1994, 1995	
Demacratic republic of Congo	2005, 2009, 2011, 2012	2005, 2009
Ivory Cost	1994, 1998, 1999, 2005, 2011, 2012	2005, 2011, 2012
Cameroon	1991, 1998, 2004, 2011	2004, 2011
Ethiopia	1992, 1997, 2003, 2008	2003, 2008
Gabon	2012	
Ghana	1988, 1993, 1994, 1998, 1999, 2003, 2008, 2014, 2016	1988, 2003, 2016
Gambia	2013	2013
Guinea	1999, 2005, 2014	
Kenya	19988, 1989, 1993, 1998, 2003, 2008, 2009, 2014, 2015	2003, 2008, 2009
Comoros	1996, 2012	
Liberia	1986, 2006, 2007, 2008, 2009, 2011, 2013, 2016	2011, 2013
Losotho	2004, 2005, 2009, 2010, 2014	2014
Madagascar	1992, 1997, 2003, 2004, 2008, 2009, 2011, 2013, 2016	
Mali	1987,1995,1996, 2001, 2006, 2012, 2013, 2015	2006
Malawi	1992, 2000, 2004, 2005, 2010, 2012, 2013/2014, 2015, 2016, 2017	2000, 2004, 2010, 2012, 2014, 2015, 2017
Mozambique	1997, 2003, 2004, 2009, 2011, 2015	2009
Nigeria	1990, 2003, 2008, 2010, 2013, 2015	
Niger	1992, 1998, 2006, 2012	2006
Namibia	1992, 2000, 2006, 2007, 2013	2013
Nigeria (Ondo state)	1986, 1987	
Rwanda	1992, 2000, 2005, 2007, 2008, 2010, 2011, 2013, 2014, 2015, 2017	2010, 2011, 2013, 2014, 2015, 2017
Sudan	1989, 1990	
Sierra Leone	2008, 2013, 2016	2008, 2013
Senegal	1986, 1992, 1993, 1997, 2005, 2006, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017	2005, 2010, 2012, 2015, 2016, 2017
Sao Tome and Principe	2008, 2009	
swaziland	2006, 2007	2006, 2007
Tchad	1996, 1997, 2004, 2014, 2015	
togo	1988, 1998, 2013, 2014, 2017	1988, 2013, 2014
Tanzania	1991, 1992, 1996, 1999, 2003, 2004, 2005, 2007, 2008, 2009, 2010, 2011, 2012, 2015, 2016, 2017	2003, 2007, 2010, 2012
Uganda	1988/1989, 1995/1996, 2000, 2001, 2006, 2009, 2010, 2011, 2014, 2015, 2016	2000, 2009, 2011, 2014, 2016
South Africa	1998, 2016	
Zambia	1992, 1996, 1997, 2001, 2002, 2007, 2013, 2014	2007, 2013, 2014
Zimbabwe	1988/1989, 1994, 1999, 2005/2006, 2010/2011, 2015	1999, 2005, 2010, 2015

Table A.2: Son Bias and HIV: Heterogeneity by Age at First Birth (***Fathers***)

Dependent variable: HIV: (0/1)

Panel A: fathers whose 1st birst was between 15 to 24 years old

VARIABLES	(1)	(2)	(3)	(4)
first child is girl	-0.00110			
	(0.00268)			
propensity to have more daughters		0.00265		
		(0.00274)		
has only daughters			0.00953**	
			(0.00434)	
actual son percentage smaller than ideal son percentage				0.000846
				(0.00268)
Observations	38,145	39,177	38,145	39,184
R-squared	0.097	0.098	0.098	0.098

Panel B: all fathers whose 1st birst was between 25 to 39 years old

VARIABLES	(1)	(2)	(3)	(4)
first child is girl	-0.000955			
	(0.00267)			
propensity to hav emore daughters		-0.00147		
		(0.00273)		
has only daughters			0.00484	
			(0.00431)	
actual son percentage smaller than ideal son percentage				-0.00496*
				(0.00269)
Observations	34,837	35,561	34,837	35,569
R-squared	0.096	0.096	0.096	0.096

Panel C: all fathers whose 1st birst occurred at the earliest at 40 years old

VARIABLES	(1)	(2)	(3)	(4)
first child is girl	-0.00867			
	(0.00618)			
propensity to hav emore daughters		-0.00254		
		(0.00626)		
has only daughters			-0.0130	
			(0.00875)	
actual son percentage smaller than ideal son percentage				0.00847
				(0.00651)
Observations	6,683	10,492	6,683	10,642
R-squared	0.055	0.040	0.055	0.038
Age	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓
Education	✓	✓	✓	✓
Wealth	✓	✓	✓	✓
Religion	✓	✓	✓	✓
Ethnic homeland by Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the ethnic group level. $*** \ p < 0.01$, $*** * * p < 0.05, *p < 0.1$. *The dependent variable is a continuous variable representing the lifetime number of sexual partners of mothers.*

Table A.3: Tribes Statistics

Variables	Obs	Mean	Std.Dev	Min	Max
	Bride price is optional or tokened				
Matrilineal kinship	34,342	0.088201	0.2835912	0	1
Patrilineal kinship	34,342	0.7179838	0.4499877	0	1
	Bride price mandatorily paid to bride's family				
Matrilineal kinship	324,598	0.0814854	0.2735795	0	1
Patrilineal kinship	324,598	0.6823301	0.465571	0	1

Chapter 2

The Digital Road to Inclusive Finance in Sub-Saharan Africa

2.1 Introduction

Small and medium-sized businesses are the driving force behind job creation in Sub-Saharan Africa. They account for over ninety five percent of businesses (Quartey et al. (2017), Fjose & Grunfeld (2010)), and represent over sixty percent of employment in the private sector (Ayyagari et al. (2011), Ayyagari et al. (2007)). Despite playing an essential role in economic development, these enterprises face significant financial challenges. They have great difficulties to obtain financing from banks and financial institutions, in spite of the fact that the latter generally have significant excess liquidity (Pontes et al. (2021), Khemraj (2008), Saxegaard (2006), Agenor et al. (2004)). In 2020, for example, the average broad money (M2)[1] to the GDP ratio was 39.2 for sub-Saharan African countries, compared to 62.8 for Europe and central Asia, 77.9 for Latin America and Caribbean, and 111.3 for North American countries. These Statistics reveal a low implication of the financial system in private sector's investments in sub-Saharan African countries, which slows the growth and development of these economies.

The reluctance of banks to finance the private sector in sub-Saharan Africa is generally explained by informational asymmetry on the ability of potential borrowers to repay loans, and also by banks' inability to reach solvent borrowers. The literature identifies several obstacles to small and medium-sized enterprises' access to bank financing: information asymmetry (Pretorius & Shaw (2004), Kitindi et al. (2007), Stiglitz & Weiss (1981); networking and information asymmetry (Le & Nguyen (2009), Kolehmainen et al. (2007), Shane & Cable (2002), Owualah (2002)); absence of collateral (Elali et al. (2013), Kira & He (2012), Barbosa & Moraes (2003), Coco (2000)); and education, abilities, and managerial performances

[1] The monetary aggregates M1, M2, and M3 make up the money supply M. These totals are the accounting representation of supply-valued money. They gauge the degree of economic agents' liquidity. M1 is made up of demand deposit funds, fiduciary money, and the coins and bills in current accounts can be used as means of payment. M2 is equal to M1 plus all deposits and short-term loans. Central banks can affect the M2 mass by increasing their economic stimulus (or encouraging consumers to limit their spending by decreasing this liquidity). By changing its policy interest rate, a central bank can achieve this. However, because commercial banks in sub-Saharan Africa are in a situation of excess liquidity, the interest rate is no longer a valuable tool for an efficient monetary policy in this region. M3 is a more expansive total that consists of M1 and M2 plus highly liquid long-term deposits and loans (over two years). Commercial banks in sub-Saharan Africa have excessive liquidity because they choose to maintain sizable reserves within their Central Banks. This implicitly means that they are providing comparatively little credit to the private sector (i.e., M2 and M3). Since Central Banks utilize the money supply as a tool to govern the financial system and the economy in order to foster growth and keep inflation under control, and given that the impacts attributable to variations in this level of liquidity are taken into account while determining their monetary policy, Central Banks will be unable to efficiently employ the monetary policy instrument to encourage economic growth in sub-Saharan Africa, as a result of that excessive liquidity reserve of commercial banks.

(Hisrich & Drnovsek (2002), von Broembsen et al. (2005), Smallbone & Welter (2001), Martin & Staines (1994)). With the growing development of information technologies, some of these barriers are likely to be removed, especially in developing contexts. This hypothesis motivates an empirical investigation of the question of how access to financing responds to the expansion of information and communication technologies.

In this paper, we analyze the causal effect of high-speed Internet on access to bank credit in Africa. Fast internet could help reduce financial market inefficiency related to information asymmetry, thereby improving the possibilities of obtaining a bank financing. In addition, high speed internet, by promoting job creation and the productivity of individuals and firms (Hjort & Poulsen, 2019b), increases individuals' chances of obtaining collateral to guarantee a loan. Similarly, access to fast Internet is likely to increase financial literacy and skills, facilitating access to information necessary for successful loan applications. The data seem to corroborate these theoretical explanations. Indeed, despite the fact that monetary credit to the private sector is still low in Africa, it has increased over the last two decades (from around 24% of GDP in 2002 to 28% of GDP in 2020), mirroring significant expansion in Internet infrastructure and increase in the number of Internet users (see Figure 2-12). We also note that while the number of secured Internet servers positively correlates with domestic credit to the private sector, it negatively correlates with the ratio of bank liquidity to bank assets (Figure 2-12), suggesting that Internet expansion in Africa over the last two decades has increased both banks' and businesses' ability to grant and receive loans. However, these figures do not establish a causal relationship running from access to high-speed Internet to access to bank credit.

To determine if access to high-speed Internet causally increases access to bank credit, we exploit both the gradual arrival of submarine Internet cables on the coast of some African countries and maps of the terrestrial cable networks. This information is combined with individual-level data from the Living Standard Measurement Study (LSMS) surveys, which are nationally representative surveys conducted by the World Bank Group in African countries since 1985. These surveys collect a wide range of information, including information on health, education, welfare, employment, income and finance. For our purpose, we use information on employment, income and finance. For causal identification, we exploit the fact that LSMS surveys were conducted in 4 African countries before and after the arrival of submarine Internet cables in these countries. We therefore apply a difference-in-differences framework to this setting to identify the causal effect of access to submarine Internet on access to bank credit. Our research design is similar to that in Hjort & Poulsen (2019b), but we address a different question.

We find that the advent of submarine Internet has brought financial institutions closer to communities and facilitated the creation of financing opportunities for individuals and firms. Specifically, exposure to submarine Internet increases the likelihood that a location or a community hosts a financial institution by 4.56 percentage points. This effect is substantial and robust to the inclusion of a large number of control variables. We also find that connection to submarine Internet increases the probability of a business owner obtaining loans from a bank by 5.13 percent.

In addition, we explore how access to fast Internet affects perception of the quality and availability of bank financing products.We find that the perception that non-agricultural credit is more accessible has increased by 7.77 percentage points in response to access to high speed Internet. Similarly, access to high speed Internet has increased the perception that agricultural credit is more accessible has increased by 4.13 percentage points. However, in rural areas, fast internet reduces accessibility to agricultural credit by 13.1 percentage points. This latter finding, which seems counter-intuitive, is explained quite simply by the fact that the arrival of the fast Internet induces a migration of employment from the agricultural sector to other sectors of activity (as also shown in Hjort & Poulsen (2019b)).

The results obtained using individual-level data are very consistent with those obtained using firm-level data and country-level data. In particular, the country-level analysis relies on a panel of 32 countries covering the period from 2005 to 2019. We find that the arrival of submarine Internet significantly increases the number of commercial bank branches in a country. In addition, a country offers 2.8 percent more domestic credit to the private sector (% of GDP) following connection to submarine Internet. The implication of the latter result is a decrease in the ratio of banks' liquidity reserves to their assets by 2.4 points.

We document significant heterogeneities in the causal effects of submarine by place of residence. The positive impact on the presence of a bank in a community is much larger for rural areas. However, the positive effect on the probability of obtaining a loan is much larger in urban areas. These findings imply that while Internet has helped the expansion of banks more in remote locations, it has not had a comparable effect on their ability to finance economic activities in these less developed areas. In particular, we find that the effect of submarine Internet on access to agricultural credit is negative in rural areas while being positive in urban areas. Also, while submarine Internet has a positive effect on access to non-agricultural credit in both areas, its impact is larger in urban areas.

We examine possible mechanisms through which submarine Internet increases credit access for individuals and entrepreneurs. First, we gather data using Google trends to show that fast Internet access increases searches of keywords related to credit access. This suggests that this technology facilitates access to information needed for a successful loan application. Second, we show that access to fast Internet positively affects entrepreneurship, job creation, and employment. This implies that fast Internet makes it easier for individuals and firms to acquire assets (real estate, materials and equipment, etc.) that can be used as collateral in a loan. This in turn would reduce the reluctance of banks to grant loans. Indeed, using our firm-level data, we show directly that fast Internet increases the probability of a firm to acquire assets that can be used as a collateral by 3.13 percentage points. Third, we find that access to fast Internet increases the probability of a borrower to repay their loan by over 1.73 percentage points. Given that good creditworthiness sends a positive signal to banks in regards to loans repayment, this also should increase incentives for banks to grant credits.

To our best knowledge, this study is the first to show the causal effect of Internet access on credit access for individuals and firms, and on bank excess liquidity as a corollary. Our analysis contributes to the recent literature on the economic impacts of communication technologies. Studies have shown the effect of Internet access on employment (Hjort & Poulsen (2019b), Rollo & Paunov (2015)), human capital (Vasileiadou & Vliegenthart (2009), Bozeman & Rogers (2002)), economic growth and development (Espinoza et al. (2020), Wagner & Cockburn (2010), Batjargal (2007), Cockburn & Wagner (2007)), on financial and banking markets (Zhang & Zhang (2015) Gong & Liu (2016), Shen & Chen (2018)), and on health (Abbasi & Pongou (2022)). Rollo & Paunov (2015) and Hjort & Poulsen (2019b) find that enterprises that own investments in Internet technologies are more productive and more innovative than those that do not, generating an increase in new market entry and facilitating job creation. Other studies have looked at the impact of the internet on financial markets. Zhang & Zhang (2015) identify the consequences of the influx of uninformed traders into financial markets, due to internet access. Their results show that these uninformed traders adopt feedback methods and cannot follow the strategies of firms with great experience in trading, and therefore cannot affect the equilibrium of the financial market. Shen & Chen (2018) provide an overview of the role played by the utilization of digital technology and media in financial markets. They find an effect of social media on market dynamics and stocks' prices. Our study differs in its research question, scopes and analyses. We document the causal effect of fast Internet on different types of credits, document some underlying mechanisms, and derive implications for bank excess liquidity.

In the next section, we describe the conceptual framework that justifies the relationship between fast internet and credit access. In Section 3, we present the context of sub-Saharan Africa in relation to credit access. In section 4, we present the different datasets we use. In Section 5, we present our identification strategy. Section 6 shows our main findings, and Section 7 documents possible mechanisms underlying the effect of submarine Internet on credit access. Section 8 concludes.

2.2 Fast Internet and Access to Bank Financing - Conceptual Framework

Small and medium-sized enterprises (SMEs) create significant employment prospects to individuals all around the world, and are therefore increasingly viewed as an important engine of economic growth (Quartey et al. (2017), Fjose & Grunfeld (2010), Ayyagari et al. (2011), Ayyagari et al. (2007)). In sub-Saharan-Africa, SMEs represented about 95 percent of all enterprises in the private sector in 2019. Additionally, SMEs contribute to more than 70 percent of employment[2], which represents approximately 60 percent of all private sector employed labour force. With an underdeveloped financial system in that part of the world, bank credit is the principal external financing sources for SMEs.

Nevertheless, SMEs are every so often young businesses and hypothetically incapable of offering adequate collateral (Berger & Udell (1998), Beck (2013), Berger. & Udell. (2005), Beck et al. (2011)). Consequently, they are more likely to be considered as risky borrowers, which makes banks reluctant to grant them loans. Because SMEs are the essence of economic activity in countries around the world and especially in sub-Saharan Africa, understanding how fast Internet could affect their access to bank financing is important. The hypotheses presented in this section, and which will guide the empirical investigations carried out in this chapter, are all summarized in Figure (2-1).

2.2.1 Fast Internet, Opacity of Information and Banks' Willingness to Grant Loans

The economic literature demonstrates the central place of access to information in the process of convergence towards an efficient market equilibrium (Desgranges (2000), Vives (2014), Andersen & Hviid (1990), Grossman & Stiglitz (1980), Kihlstrom & Mirman (1975)). For there to be an efficient equilibrium in the lending market, the availability of information and transparency are of great importance, both for loans applicants and their suppliers. Another body of works shows that the internet, by reducing information asymmetry, has improved investor's access to information (Gajewski & Li (2015), Duarte & Young (2009), Trabelsi et al. (2008), Hodge et al. (2003), Ettredge et al. (2002)). A direct implication resulting from this body of works would be that Internet could reduce the cost of capital. Indeed, Internet, by allowing companies to make their financial and accounting information publicly available, also improves transparency. As a result, Internet can be a source of information for investors, and for financial institutions and banks.

[2] 2019 report of the International Labour Organization

Internet access improves access to information both for financial institutions (or banks) and credit applicants (individuals and companies) (Storm et al. (2017), Arnold & Beckman (2013), Bostrom & Sandberg (2009)). Absence of information on a potential borrower's ability to reimburse a loan, or a lack of transparency on how they will manage or use a loan that could be granted to them, is a significant constraint on any financial institution's decision to grant a loan. Likewise, a company in need of financing will seek for information in the financial market which will enable it to obtain the cheapest and most interesting financing. The unavailability of such market information could either lead to an absence of an equilibrium or result in an equilibrium that is inefficient. This in turn would limit the growth of both financial institutions and businesses.

As the relationship between access to information and access to financial credit has been amply demonstrated in the existing literature (Le & Nguyen (2009), Kitindi et al. (2007), Kolehmainen et al. (2007), Pretorius & Shaw (2004), Shane & Cable (2002), Owualah (2002), Stiglitz & Weiss (1981)), we will not be investigating it further. Instead, in this section, we are setting out to demonstrate how the advent of submarine internet facilitates access to the kind of information that reduces the reluctance of banks to finance the economy. More precisely, we endeavor to demonstrate how access to internet on the one hand facilitates accessibility to information; and on the other hand, we show how the information obtained promote a financing applicant's odds to acquire a loan from a financial institution.

With internet connectivity, it becomes easier to learn about the different possibilities to obtain bank financing. Also, the wide range of materials available on internet helps in the choice of the financial product that best meets one's needs. Thus, the reluctance of banks to grant loans is not the consequence of a lack of information by the loan applicant, but rather a lack of information by the entity in possession of the capacity to fund. This is the reason why we will only focus on accessibility to information for banks and financial institutions. We demonstrate how, the connectivity to high speed internet facilitates banks' accessibility to information on their potential borrowing clients. In other words, we will show how submarine internet access can enable banks in sub-Saharan Africa to get closer to their clients, and so enable them to obtain the right information more efficiently and with better accuracy regarding their solvency.

A bank that receives a financing application from a company will base its decision to grant that financing on the information it manages to gather on that enterprise; namely, the sector of activity in which it operates, its reputation in this specific market, its solvency and other accounting and financial information, such as its turnover or the value of its movable and real estate assets; as well as that of their liabilities. Access to information of this nature has been partly made possible thanks to internet. Indeed, since the arrival of the internet in sub-Saharan Africa, social communication networks, specialized in business have been created[3]. These networks catalog and provide reliable and detailed information on existing businesses in the economy, and information on their assets, liabilities, and turnover. In addition, these social networks provide a great deal of information about the attraction that the company has generated. They also provide statistics on how the public interest in that company has evolved. Once more, the comments and appraisals that the market attributes to this company and that we will find on the internet pages constitute elements that send back a revealing signal as to the future of the said company and the possibility for the bank to recover its capital invested in that company. In other words, the traces this company leaves on social networks (thanks to internet accessibility) allow it to build its reputation. This reputation partly provides information on the company's credibility and creditworthiness and allows banks to make informed decisions on loans applications. At the same time, a lack of visibility or absence of a company on internet networks could send a negative signal and thereby increase the reluctance of a bank to approve the loan request.

Another way fast internet could make it easier for financial institutions to access information about applicants for funding is through advertisements. Many companies, since the advent of the internet, have created a website. Those websites are places where information about their capital, turnover, financial and non-financial assets are made available to the public. Thus, the internet and social networks, being used as an advertising tool for these companies, are also source of information for potential investors. These internet pages thus guarantee a certain transparency of information. Thus, fast internet connectivity, by facilitating the availability and transparency of information on loan applicants, could decrease the reluctance of financial institutions to finance the economy.

The literature on bank excess liquidity also argues that the inability of banks to follow up, in order to ensure that the funds granted are used for the declared purposes, is one of the reasons why they are overly cautious about granting loans, especially when the investment constitutes the collateral. The arrival of the internet perhaps does not completely remove this barrier, but it greatly limits this obstacle. Thus, another benefit of the internet in its support of access to information about funding applicants is through mobile telephony. With the arrival of the internet, the use of cell phones has amplified, and it has become easier to trace a phone and locate the geographical position of a mobile phone user, so that an individual who has received bank financing will have less incentive to adopt deviant

[3]
https://afriquebusiness.info/
https://startups.sekou.org/
https://www.goafricaonline.com/annuaire/accessoires-mode
https://www.yasdg.com/

behavior. Connectivity to telecommunication networks makes post-borrowing monitoring possible or at least reduces obstacles to post-borrowing monitoring; this implies a reduction in the reluctance of banks to grant loans.

To summarize, a better accessibility of financial institutions regarding information on the financial and accounting health of small and medium-sized enterprises would increase the propensity of banks to grant credit. The advent of submarine internet is therefore a factor that could reduce the opacity of information on the financing market because access to the internet allows companies, for example, to disclose information on their webpages by improving their reputation, this would attract potential investors (Ettredge et al. (2002)). In addition, because Internet facilitates data collection, banks and other economic agents with excess liquidity can easily obtain information on the finances of companies, which increases the propensity of enterprises with good visibility to obtain financing. Thus, the Internet could be beneficial for both the banking and financial institutions, as well as small and medium-sized businesses in sub-Saharan Africa.

Moreover, by reducing agency and communication costs, Internet allows financial institutions to advertise their products and services to potential clients. This in turn could make it possible for potential borrowers to be better informed about the availability of financing formulas appropriate to their situation, reducing the rate of rejection of financing requests; and thereby, the reluctance of banks for financing.

2.2.2 Fast Internet, Productivity and Banks Willingness to Grant Loans

The arrival of the fast internet, through the access to information that is facilitated, could boost individual productivity and even that of businesses. Indeed, accessibility to internet makes easier the availability of information, and thereby eases the diffusion of knowledge among industries to enhance their innovation abilities, their productivity and their performance (Rollo & Paunov (2015), Hjort & Poulsen (2019b)). Rollo & Paunov (2015) and Hjort & Poulsen (2019b) have shown that the use of internet by firms positively impacts their average productivity. A higher productivity means greater income (Strain (2019)). This increase in firms' productivity therefore increases their ability to offer collaterals in exchange for bank credits, reducing the reluctance of commercial banks to grant them financing.

Furthermore, advancement in information technology (IT) is closely related to innovation in new products and services. Indeed, firms can take advantage of improvement in IT to create new markets and expand their supply. Thus, technological development, by promoting the creation of new markets for companies, improves their profits and therefore their likelihood of obtaining bank financing. The arrival of high-speed Internet may allow the expansion of new markets, not only for firms in need of financing, but also for financial institutions (through, for instance, mobile banking, with online credit application system). To put it briefly, access to the fast Internet, by increasing productivity and income, would allow small and medium-sized enterprises in sub-Saharan Africa to build up collaterals, facilitating credit access.

2.2.3 Fast Internet, Employment/Entrepreneurship and Willingness to Grant Loans

Another channel through which fast internet could impact the access to bank financing is by increasing employment. The literature on infrastructures documents the effects of infrastructures on entrepreneurship and the creation of new employment opportunities (Espinoza et al. (2020), Hjort & Poulsen (2019b), Rollo & Paunov (2015), Reuber & Fischer (2011), Batjargal (2007)). The advent of submarine internet, by promoting employment (Hjort & Poulsen (2019b)), would strengthen the possibility of borrowers to have collateral that can guarantee their loans, so that SMEs' chances of obtaining credit from commercial banks would increase. Following this rationale, we will empirically study: (1) the effects of the arrival of submarine Internet on entrepreneurship and on employment opportunities; and (2) the effects of entrepreneurship and employment on the likelihood to receive a loan approval.

2.2.4 Fast Internet, Creditworthiness and Banks' Willingness to Grant Loans

Internet expansion may also encourage firms to improve their creditworthiness. Improvement in the creditworthiness of borrowers as a result of Internet expansion stems from the fact that fast Internet has increased transparency on credit and solvency records and on bankruptcy history. Knowing that a good credit rating is a fairly good indicator about the quality of the borrower, we can logically expect that the arrival of the fast Internet, by facilitating access to information on the solvency of companies, could encourage the latter to work to the best of their ability to have good credit rating. In other words, Internet provides an incentive for firms to improve their services.

Given that a positive credit or solvency history boosts the propensity of financial institutions to provide financing, we will test the hypothesis that the repayment rate of bank loans is positively affected by fast Internet.

2.3 Banking Markets and SMEs' Credit Access in Sub-Saharan Africa

The banking market is an important component of the financial system of sub-Saharan African countries. The poor performance of this market may partly explain the slow economic growth of these countries. Indeed, theories of financial intermediation hold that the expansion of the financial market enhances the efficiency of capital and the savings rate, which positively impacts the growth rate (Christensen (2006), Levine (2004), Allen N. et al. (2001), Gelbard & Leite (1999), Bencivenga & Smith (1991), Greenwood & Jovanovic (1990), Shaw (1973), McKinnon. (1973)). The relationship between financial system and economic growth works in an important way by disabling some outer financing restrictions that might otherwise hamper development and the expansion of enterprises. Another body of works explains the poor performance of sub-Saharan Africa by relying on the demand side. This literature finds the needs for financial services and products that arise from economic development force financial institutions and banks to adapt in order to respond to market demand (Khalifa Al-Yousif (2002), Greenwood & Smith (1997), Demetriades & Hussein (1996), Ireland (1994), Robinson (1979), Friedman & Schwartz (1963)).

Despite the multiple financial reforms and the stable economic growth, banking systems in sub-Saharan African countries are nevertheless underdeveloped, with small and ineffective intermediation and reduced competition (Jean-Pierre Briffaut & Swaray (2018), Gakunu (2007)). Barriers to the development of the banking sector include small domestic markets, low levels of income, and weak creditor civil rights (Galindo & Micco (2016), Beck & Demirguc-Kunt (2006)). Consequently, for both individuals and firms, access to financing remains among the lowest in the world. In other words, sub-Saharan Africa's financial systems fall behind the rest of the world in term of financial penetration.

Despite this developmental delay in the financial sector, sub-Saharan Africa has been on a low, but continual expansion path for the last twenty years (Mlachila et al. (2016)). This constant movement represents a substantial change, compared to preceding decades. These positive advancements can be credited to many elements, such as continuous improvements in policies, debt relief, substantial natural resources, and expanded investments. Thus, the resilience of sub-Saharan African countries has increased, as witnessed by the modest impact of the recent global economic crisis in the region. However, the prerequisites for social and economic expansion diverge considerably within regions, and this has added to the irregular speed of growth across the sub-Saharan African region.

From the last two decades, the financial systems of many sub-Saharan African countries have added significantly to the development and productivity of their banking markets

(Mlachila et al. (2016)). As a result, the capitals of commercial banks have been reinforced with better practices of risk management. Consequently, credit to the private sector has increased, although from a low base; and the vast majority of banking systems in sub-Saharan Africa have demonstrated significant resilience in regards to events related to the recent international financial crisis[4] (F. Allen & Giovannetti (2011)). The causes of this improvement in the provision of financial products and services, however, remain largely undocumented. We will show that Internet expansion has significantly contributed to this positive outcome in Africa.

The issue of credit access for small and medium-sized enterprises is a matter of great relevance in many sub-Saharan African countries for at least two reasons. First, the vast majority of these countries have their economies based primarily on the activity of these enterprises (Quartey et al. (2017), Fjose & Grunfeld (2010), Ayyagari et al. (2011), Ayyagari et al. (2007)). In fact, with very low employment opportunities in those countries, over 60 percent of the working population operates in the tertiary sector of small and medium-sized businesses. However, economic theory teaches us that when the population grows faster than economic growth, we inevitably end up in a Malthusian trap characterized by some high level of unemployment. In the case of sub-Saharan Africa, where the population is growing quite quickly, but with a financial system that does not allow small and medium-sized enterprises to expand, the long-term consequence would be an increase in the unemployed labor force, and therefore an increase in poverty.

Secondly, financing accessibility, being one of the main barriers to the activity and growth of sub-Saharan African enterprises, financing possibility is one barrier among many other obstacles encountered by entrepreneurs throughout nations (Idrees et al. (2020), Kazimoto (2014), Kira (2013), Okpara (2011), Irwin & Scott (2010), Balling et al. (2009)). Business expansion is mostly obstructed by a lack of infrastructures. A characteristic example is that of the accessibility to main road, poor means of transportation, and poor access to technology and information (Idrees et al. (2020), Cusolito et al. (2016), Kazimoto (2014), Ponmani (2011), Okpara (2011), Southiseng et al. (2008)). Indeed, the landscape of sub-Saharan Africa shows a significant infrastructure deficit. Therefore, if economies of this region want to catch up in infrastructure development, innovations and improvements in financing patterns are required to be able to finance investments in infrastructures.

The existing literature on financing and economic growth suggests that monetary policies and central banks' main instruments cannot influence economic activity in developing countries (Starr (2005), Greenwood & Jovanovic (1990)). In these countries, the low level of development of the financial system, through its low market competition and its excess

[4] 2009 Report of the African Development Bank

liquidity reflects the inefficiency of financial markets to finance economic activities. Bank excess liquidity deviates monetary transmission mechanisms, decreasing the efficiency of the interest rate, which is the traditional monetary policy instrument (Gigineishvili (2011), Mishra et al. (2012), Greenwood & Jovanovic (1990)).

Bank excess liquidity means that commercial banks continuously keep extra reserves. That is, banks incessantly hold more deposits than the credits they allow to those in need of financing. Thereby, they retain important amount of assets which do not produce any return. This excess liquidity of banks has significant and adverse effects on economic growth. Actually, with excess liquidity, a commercial bank does not need to borrow funds from the central bank. Thereby, the regulating power of the central bank's interest rate is reduced, and can no longer be used as a powerful instrument of monetary policy. A direct implication is the decrease of the transmission effects of the monetary policy (Ganley (2004), Bathaluddin et al. (2012)).

It is hypothesized that the low competition in the financial system and the high cost of information access are key determinants of bank excess liquidity in less developed countries (Jovanovic et al. (2015)). The significant changes in the landscape for banking systems in many sub-Saharan African countries, such as the expansion of mobile banking and mobile methods of payment, facilitated by the spread of new technologies, therefore lead us to question about the impact of the internet on access to bank financing. As these countries are deemed to be in the state of excess of liquidity, we are interested in knowing whether information technology shocks, and in particular the arrival of the fast Internet, have made it possible to reduce this excess liquidity via the development of financial markets and its impact on the financing granted to small and medium-sized business.

2.4 Data

For our main analysis, we combine data on Internet optical fiber nodes, timing of arrival of submarine Internet cables, and Living Standards Measurement Study (LSMS) surveys conducted between 2005 to 2019 in four sub-Saharan African countries. We use other data sources including the World Bank Enterprise Surveys, information on electricity grids to control for access to electric power, and google trends data. These data are described below.

2.4.1 Living Standards Measurement Study (LSMS)

The Living Standard Measurement Study (LSMS) has been launched by the World Bank Group to collect longitudinal household data in some African countries, with the key intention to study and enlighten development programs. LSMS surveys are performed by a close collaboration with the national offices of statistics of its affiliated countries, which are Ethiopia, Ivory Cost, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda. For the specific case of sub-Saharan African countries, the World Bank, through the LSMS, created and implemented methods of multi-subject matter, across nations, with a focus on agriculture. The fundamental motivation of this mission is to expand development and proficiency in research on the links between agribusiness, non-agricultural activities, financial conditions, health and economic growth.

The LSMS uses a stratified two-stage sampling procedure. In the first stage, enumeration areas (EA) are selected with a probability proportional to the total number of households in the EA. In the second stage, households are selected in each EA. The draw of households in each EA is done after a listing operation whose purpose is to update the selection of households. In each household, information is collected on various demographic variables, income, consumption, employment, health, education, and financial decisions. In each subsequent survey round, only households residing in EAs selected in the first round are interviewed. Most of these households were interviewed in previous rounds, but some might have moved to other locations, therefore leaving the study. Also, some new households may enter the study.

In this paper, we analyze 18 LSMS surveys from 4 sub-Saharan African countries (see Table(A.1)). These surveys were conducted between 2005 and 2019. The frequency at which the data collection is carried out is determined, on one hand by each country, and the other hand according to the availability of funds to cover the costs. All LSMS surveys used in our analysis are listed in Table (2-2). Of the LSMS cycles available for sub-Saharan African countries, only those of Malawi, Niger, Tanzania and Uganda contain cycles before and after the arrival of the submarine Internet. We therefore carry out our analyzes on a repeated cross-sectional sample made up of all cycles in these four countries. Our sample includes individuals aged between 20 and 65 years old. Over 52 percent of these individuals are women, approximately 38 percent have a secondary level education, and 80 percent reside in rural areas. Approximately 35 percent of individuals are entrepreneurs. Of these individuals, 31 percent operate in the agricultural sector. Roughly 29 percent of individuals have a permanent employment. Additional features on this sample are shown in table (2-5).

Geo-referenced data contained in the surveys made it possible to combine information from these surveys with data on Internet nodes and electricity grids. In our final dataset, an average household resides in an area that is 9.8 km from an electricity grid and 66.6 km from an optical fiber node.

2.4.2 The World Bank Enterprise Survey (WBES)

The World Bank Enterprise Survey (WBES) is a countrywide representative sample of businesses from all sectors of activity with at least five workers. The Enterprise Surveys were conducted towards business owners and the highest supervisors of the private sector industries. The surveys include information on enterprises' sources of funding, assets, and ability to provide collateral. We exploit the WBES data from 2003 to 2018 of coastal nations that had survey cycles before and after the landing date of submarine internet. These countries are Ghana, Kenya, Mauritania, Nigeria, Senegal, and Tanzania.

After pooling observations from these countries, we obtain a sample of 15,057 firms. Around 30 percent of the firms are located in a city in which at least one terrestrial optical fibre passes through, and 75.6 percent of the firms are connected to high speed internet. Only around 17 percent of the firms have a line of credit.

2.4.3 Internet Nodes

We use data on long haul optical fibre nodes covering the period from 2009 to 2019 from Africa Bandwidth Maps (published by Hamilton). The map displays in great detail Africa's terrestrial, satellite and submarine cable transmission networks. The Hamilton's map contains: fibre optic networks (operational, under construction, planned and proposed), fibre optic networks, microwave networks (operational and planned), as well as submarine cables (operational, under construction and planned or proposed) for over 300 network operators in more than 50 African countries, and 72 submarine cable systems. The dataset includes geographic coordinates of fiber optic nodes. This allows us to compute the distance between each optical fiber nodes and the centroid of each enumeration area (EA) in the LSMS dataset. For each of the fiber nodes, we have information on the year it became effective. Figure (3-4) displays the evolution of nodes over the years.

We also use information on the landing points and the arrival dates of submarine Internet produced by Mahlknecht [2014]. We combine this information with the location of optical fiber nodes to determine the connectivity of an enumeration area (EA) to the Internet and to submarine (or high-speed) internet. We consider an enumeration area as connected to Internet when the distance between the centroid of that EA and the location of the closest optical fiber node is at most 2 kilometers. Approximately 4.4 percent of individuals in our main sample are connected to the internet. In a survey, an EA is connected to submarine internet if that EA is connected to a fiber node and if the survey was conducted after the arrival of submarine Internet cables in the country. Around 3 percent of individuals in our sample are connected to submarine internet.

2.4.4 Electricity Grids

Arderne et al. (2020) introduced the first synthesis map of the global power grid obtained by compiling available open source data (Gridfinder and OpenStreetMap data), using information on the location of electric networks. The open-source tool "Gridfinder" is exploited to forecast the geographic position of electrical networks using satellite broadcasting of nocturnal light. A specific set of rules are used to calculate the most probable geo-location to produce electrical light. Then, these target-locations (sources of electrical lightning) are linked to actual power networks through a least-cost path algorithm. The resulting OpenStreetMap data shows that 97 percent of the global population lives within 10 km of electricity lines of voltage greater than 10 Kilo-volts, while the vast majority of those over 10 kilometers are across Sub-Saharan Africa.

We combine our LSMS dataset with the global power grid map produced by Arderne et al. (2020), and we compute the distance to the nearest power system. Our dataset reveals that the average distance to the nearest electric grid is 9.81 kilometers, with 226 kilometers being the maximum distance. An EA is consider to be connected to electrical power if the distance between the centroid of that EA and the closest electricity grid is less than 10 kilometers.[5] Nearly 25 percent of individuals in our final dataset are connected to electrical power.

We define grid cells of 20 by 20 kilometer distance and calculate the distance between the centroid of each grid cell and the closest electrical power source. Each bin is either connected to electricity or is not connected. When a bin is located within 10 kilometers of the closest electricity grid, it is considered as connected to electricity. We control for the connectivity of the grids to the electrical network as fixed effects.

2.4.5 Google Trends Data

The main database on which this study is based does not allow us to empirically demonstrate how the advent of submarine internet facilitated access to financial information. However, we use the data provided by Google Trends in order to assess the relationship between high-speed internet and access to useful information for the financing market. Known as Google Trends, the website measures the popularity of top words search within several areas. Google trends displays the trend of word search with graphs that show the popularity of top search interrogations. In the area that concerns us, we want to know whether individuals use the internet tool to obtain information that will be useful to them in order to obtain funding or to make a good investment. Indeed, a bank, knowing the credit score of a loan applicant,

[5]The average distance from the electrification grids being about 10 kilometers, allows us to justify such a choice.

will take decision to grant financing or not, based on the credit score of those applicants. We will thereby, focus on the popularity of the word "credit score".

Google Trend gives monthly search frequencies for the term "credit score", by country. We would have used this information for the 7 countries included in our study. But, due to a large amount of missing data for these countries, we will use that of South Africa, which is fortunately a country in sub-Saharan Africa.

Since Google Trends provides monthly data and being interested in the period from 2004 to 2021 (that is a total of 215 observations), the size of the sample required to estimate a parameter at a confidence level of 95 percent would therefore be at least 138 observations, according to Rea (1997).[6] Indeed, Table (2-1) gives a general picture of the available data. The only country in sub-Saharan Africa meeting this requirement is South Africa. Therefore, to use a statistically representative sample, we will rely only on data from South Africa.

Table 2-1: Statistics on Google search for "Credit Score"

Countries	Observations	Missing observations	Percentage of missing values(%)
Canada	215	0	0
South Africa	188	27	12.6
Nigeria	121	94	43.72
Ethiopia	40	175	81.4
Malawi	13	202	94
Mali	0	215	100
Niger	7	208	96.7
Tanzania	39	176	81.9
Uganda	37	178	82.8

Additionally, we compare South Africa to Canada to determine if, following the introduction of fast Internet in South Africa, the search of "credit score" in this country has converged to that of Canada where fast Internet was introduced earlier. This comparison is pertinent for several reasons. South Africa is an economy that stands out among African countries for its performance in terms of economic development and innovation and has more than one similarity with Canada; technological indexes quite close around 4 to 5; interest rates

[6]According to Rea (1997), the size of a statistically representative sample allowing statistically valid estimates to be computed with a 5 percent margin of error would be obtained from the following formula :

$$n = \frac{t_p^2 * p(1-p) * N}{t_p^2 * p(1-p) + (N-1) * y^2}$$

, where n is the sample size, N the target population size, and p the expected proportion of the population, generally set to 0.5 by default. The sampling confidence interval is t_p and margin of sampling error is y (t_p=1.96 for y=5%).

spread, defined as the Difference between the lending interest rate and the savings interest rate, around 4 percent, and economic growth comparable to -3.7 for Canada and -2.83 for South Africa. In other words, these two economies have comparable characteristics continent wise, which allows us to make a difference-in-differences comparison type. This justifies our choice to compare the effect of the fast Internet on access to information between Canada and South Africa.

2.4.6 Dependent Variables

We analyze the causal effect of connection to submarine Internet on loans granted by commercial banks. We focus on six main outcome variables that capture both the supply and the demand sides of the lending market. We describe these variables below.

- First, it is important to know whether there is a real demand for financing from entrepreneurs. Indeed, if there is no need for financing, then the bank excess liquidity in sub-Saharan African financial institutions becomes totally devoid of interest, and therefore all the debates around the excess liquidity of African banks curbed by a lack of financing would become irrelevant. Therefore, first of all, we are interested in knowing if there are any companies or entrepreneur in the need for bank financing.

- Secondly, once we have verified that a real demand exists, we are interested in knowing whether the possibility of bank financing is possible. To do this, we exploit a dichotomous variable which takes the value 1 if there is a bank in the community, and the value 0 otherwise.

- Once we have verified that there is indeed, a demand (the need for financing), and an offer, we are interested in knowing whether there is a certain equilibrium in the banking market allowing to meet this demand for funding; or at least, if the advent of the submarine internet has positively impacted the way financial institutions react to entrepreneurs' (and businesses') demand for loans. Therefore, we are using a variable which will take the value 1 when the source of financing was a bank loan, and the value 0 if not.

-The last 3 variables permit to be better informed about the progress of the quality of the accessibility to financial and banking services since the arrival of the fast internet. They take the value 1 for a better accessibility and the value 0 when the respondent's evaluation does not indicate a better quality.

To evaluate the impact of fast internet on the banks' reluctance to grant loans, we explore its effects on several outcomes:

- We test its effect on the demand for bank financing. This variable takes the value 1 if a business had requested for a bank financing and the value 0 if not.
- We also look at its effect on the availability of a commercial bank in a community. Indeed, the LSMS provides us with the information on whether a community has a commercial bank.

The variable takes the value 1 when it is the case and 0 otherwise.
- Another test we perform is the evaluation of its impact on the supply of bank financing. The outcome variable we explore is a dummy variable indicating whether an individual or a firm has received a bank credit.
- The last set of individual level outcomes is related to the appreciation that individuals and firms have towards the accessibility to credit. Those variables take the value one when the entities think that the accessibility to financing products is improved.
- We also explore the effects of high speed internet at the aggregate level. The outcome variables here are the number of commercial banks, the share of domestic credit to the private sector and the ratio of bank liquidity reserves to the bank asset; which are all continuous variables.

2.5 Identification Strategy

We use a generalised difference-in-differences framework with fixed effects and multiple treatment periods to identify the causal effect of connection to submarine Internet on access to loans in sub-Saharan Africa. We estimate the following equation using ordinary least squares:

$$FIN.ACC_{ijct} = \alpha_0 + \alpha_1 Fast.internet_{it} + X_{ijct}\beta + \sigma_j + \gamma_t + \lambda_c + \epsilon_{ijct} \qquad (2.1)$$

Where $FIN.ACC_{ijct}$ is the outcome variable for an individual i in grid cell j, country c and at time t. In this specification, the parameter α_1 measures the effect of the arrival of submarine internet on access to a loan granted by a commercial bank. We defined the connectivity to internet based on the distance between an individual's location and the closest fiber node. Individuals are identified as connected to internet if they are located at a maximum of 2 kilometers from the closest optical fiber node. Hence, an individual is considered as treated (or connected to fast-internet) if they live within 2 kilometres of a fiber node and there is a backbone network in their country connected to at least one submarine cable.

A positive parameter (α_1) means than submarine internet improves the ability of an individual or a firm to obtain a loan or improves financial institutions' willingness to grant loans to businesses. In other words, a positive value of our main coefficient means that submarine internet has caused bank to better finance economic activity in African countries. X_{ijct} is a vector of controls comprising age at the survey, place of residence, and gender when we use the LSMS data.

The terms σ_{ij}[7], γ_t and λ_c capture the grid cells j (where individual i is located) connectivity fixed effect, year fixed effects and country fixed effects, respectively.

Correcting for year fixed effects allows to consider all the time-variant factors common to all grid cells, that may be correlated with the construction of internet fiber nodes or submarine cables and the granting of loans by financial institutions. We assume that the error term, ϵ_{ijct}, is not correlated with any of the variables that measure the accessibility to submarine internet, nor the other baseline controls.

How is Internet Connectivity Defined?

Recent years have seen a significant increase in the deployment of optical fiber in Sub-Saharan Africa, with an increase in the network length from 400 thousand kilometers in 2009 to over a million in 2019. In terms of information transmission, fiber optics are the most effective technology. There is less attenuation or loss of signal as one moves away from the network connection nodes with this technology. With its high speed and large bandwidth, it is possible to provide high-speed Internet access to many people from the same network. The most significant advantage of fiber optic cable is the amount of information that can be transmitted per unit. Furthermore, because it has less signal degradation, it not only results in a faster Internet connection, but also in an Internet connection with a large signal transmission range, as shown in figure (2-5). It is a type of terrestrial backbone cable distinguished by the fact that it is a very high-speed electrical cable capable of transmitting signals to more than hundred kilometres.

Two types of optical fibers exist: (1) The optical fiber with a single mode of transmission is known as a single mode optical fiber. It has the property of having a very low signal attenuation, which improves signal transmission speed. This type of wire can transmit at a speed of 100 Gb at five kilometers distance. It is, however, mostly utilized for very long-distance signal transmission due to its extremely high cost. (2) The most commonly used type of cable is multimode fiber optic cable. It has the ability to route multiple modes at the same time, allowing the system to route large amounts of data at the same time. It is commonly used to transmit signals over short distances. Multimode fiber optic cables can transmit signals of 100 Mbps up to 2 km.

According to figure (2-5), the multimode optical fibers emit at a distance of 2 kilometers for a speed of 100 Mbps, and at 1.1 kilometers for a speed of 1Gb. In this study, we consider as connected to Internet those who receive a connection at a speed at least equal to 100 Mbps. Therefore, we define the limit of 2 kilometers as the threshold. People located within a radius of 2 kilometers are considered to be connected to Internet.

[7]This fixed effect allows to control for all time-invariant changes in access to financing that may possibly be associated with access to high-speed internet.

The Causal Effect of Internet on Credit Access

Figures (2-6) shows how formal credit accessibility relates to Internet access. It can be seen that as the fiber optic network expands, the volume of credit granted to the private sector as a percentage of GDP also increases. This is consistent with the negative slope of the relationship between the number of fiber optic connection points and the ratio of bank liquidity to their assets. Note that in absolute values, the slope of the curve of credit to the private sector and that of the volume of bank liquidity are quite close. It would therefore not be wrong to think that the development of the fiber optic network, by improving Internet accessibility, would reduce the reluctance of banks to finance the private sector, thus leading to a higher volume of credit. The third quadrant of this figure also presents facts consistent with the first two, showing a negative relationship between the distance to the nearest optical fiber and the volume of credits. Indeed, with this technology that has a limited emission distance, as the odds of being connected to Internet decreases, so does the accessibility to credit. We exploit those distances to estimate the effects of investments in ICT on the credit accessibility.

Several factors determine Internet connectivity within countries. Those factors may pertain to their growth and development level. Actually, telephone cables were initially the means by which sub-Saharan Africa gained access to the Internet in the early 2000s. Back then, most of the Internet network was made up of terrestrial cables built by telecommunication companies. Many limitations of these terrestrial cables caused them to be less efficient, including high installation costs, long distances travelled for cables, long transmission time and high communication costs. These cables are also at risk of being destroyed by bad weather, such as thunderstorms, lightning, or tornadoes. Apart from the fact that the broadcasting or the emission distance of this type of Internet connection is quite short, this essentially monopolistic market tended to only be occupied by large private or governmental companies able of bearing such costs. Thus, the construction and maintenance of these terrestrial cables are not only expensive, but strongly correlated with the evolution of economic activity.

As a result of submarine cables arriving in sub-Saharan Africa, many of these limitations seem to have been alleviated, since not only have installation and maintenance costs decreased, but also the time it takes for information to be delivered has been reduced. Submarine Internet landing outside of major economic centres, connected to terrestrial cables, arrived in sub-Saharan Africa for these main purposes. With the expansion of the submarine Internet network, transmission speeds and emission distances have improved, allowing better information exchange between Africa and the rest of the world for a lower cost. Figures (2-3) show the tremendous evolution of the terrestrial network over time, especially since the advent of submarine Internet (figure (2-4))

For their exogenous characteristic, the different arrival dates of submarine cables on the coasts of sub-Saharan African countries enable us to establish a causal relationship between the Internet access and credit access. The date at which the first submarine cable landed at a coast, connecting it to terrestrial Internet network is the treatment date for countries connected to that specific submarine cable. SEACOM is a submarine cable of thousand kilometers which arrived on the African coast in the third quarter of 2009. Its commissioning in 2009 simultaneously supplied Tanzania, Uganda, Mozambique, Kenya and South Africa. Thus, in our LSMS sample, we have Tanzania and Uganda which are treated in 2009; and Kenya for the WBES. The Africa Coast to Europe submarine cable (ACE) landed in December 2012 along the west coast of Africa, simultaneously connecting several coastal and non-coastal countries such as Niger, Nigeria, Ghana, and Mauritania. Thus, Niger was first treated in 2012 by receiving submarine Internet thanks to the ACE cable. However, countries such as Ghana, Nigeria, Senegal were first treated at the beginning of the third quarter of 2010, by the arrival of the MainOne cable.

We define two groups. The treated group is composed of individuals with high-speed Internet access (submarine Internet), located within 2 kilometers of the closest fiber node after the landing date of the submarine cable connecting their country. The non-treated group consists of people who live more than 2 kilometers from the nearest terrestrial fiber optic node, as well as those who were interviewed before the arrival of submarine cables.

An irrefutable identification strategy would be one that shows that in the absence of the advent of the submarine Internet, the characteristics of the treated and the control groups would have had similar trends. Given the technical impossibility of making such a demonstration, we could rather show that the two groups present parallel trends in their characteristics before the arrival of fast Internet. However, our database presenting only one year of survey before the treatment, does not allow us to appreciate the trends over time of the pre-treatment characteristics.[8] Nevertheless, we compare the characteristics of the two groups prior to submarine Internet arrival (table (2-4)). However, with WBES data, we do not have this data issue. One can appreciate the pre-treatment trend on a graph (figure (2-9)).

[8]See Hjort & Poulsen (2019a) for a similar problem. Our research design follows that of the latter study.

2.6 Average Treatment Effects of Submarine Internet

2.6.1 Financial Loans: The Demand Side

Our LSMS sample contains 168,457 individuals who had claimed to run their personal businesses, which represents 64.92 percent of the sample. Over 66 percent of these individuals reported that they applied for credit from at least one financial institution. In Panel A of table (2-6), we find that the advent of fast Internet has increased demand for bank credit among business owners (column (1)). In other words, when having access to fast Internet, business owners are 2 percentage points more likely to have applied for financing. This result reflects the fact that investment in high-speed Internet generates momentum in private sector entrepreneurial initiatives.[9] This finding is consistent with other studies that have found positive effects of investment in infrastructures on a wide variety of economic outcomes (Donaldson (2010), Jedwab & Storeygard (2017), Ejaz Ghani (2016) Allen N. et al. (2001), Shatz et al. (2011), Aschauer (1989), Okoye et al. (2019)). This enthusiasm for entrepreneurship would therefore generate a need for financing. The conjecture is that, with accessibility to submarine Internet, the number of innovative ideas that entrepreneurs have is growing, and the financing needs to carry out these investment projects therefore generate a demand for credit. We explore this assumption in more depth in the next section.

In the Columns 2 and 3 of table (2-6), we compare rural and urban entrepreneurs' credit demand. We find that Internet positively affects the credit demand in both areas, but the magnitude of the effect is higher in urban areas (5.16 percentage points) compared to rural areas (0.52 percentage points). This could be explained by the fact that the level of economic activity and innovative ideas is greater in urban areas, therefore generating a greater demand of financing in response to access to fast Internet.

The positive and significant impact of the advent of submarine Internet on entrepreneurs' demand for financing shown in Panel A of table (2-6) is robust to controlling for age, gender, year and country fixed effects, and a fixed effect for the interaction between grid cell and connection to electricity. Column (1) additionally controls for place of residence. If submarine Internet access causes the increase in the likelihood of an entrepreneur to apply for some financing, one can easily refute the idea that sub-Saharan African banks have excess liquidity because they do not receive demands for funding. Will these stimuli that the arrival of submarine Internet exerts on the credit demand also be present on the supply side? In such case, will it be greater, so that excess bank liquidity and the banks' reluctance to grant loans be reduced? We answer these questions in the next sections.

[9]Indeed, we find that access to the fast internet increases entrepreneurial initiative by increasing the propensity of individuals to have an activity that generates income (Panel C, table(2-18)).

2.6.2 Financial Loans: The Supply Side

We find that the arrival of submarine Internet has brought the banks closer to the communities. The arrival of submarine Internet has increased the probability of having a bank in a community. In Panel B of table (2-6), we find that, compared to communities not connected to submarine Internet, communities connected to this technology are 4.56 percentage points more likely to have at least one bank in their locality (Column (1)). This result is consistent with the analysis performed at the country level, where we find that countries connected to fast internet have on average 1 to 2 more commercial bank branches, compared to those not connected to submarine internet (table (2-15)).

More interestingly, we find that the arrival of fast Internet is more beneficial to the establishment of financial institutions in rural areas than in urban areas. Indeed, from Columns (2) and (3) of table (2-6), we can see that access to submarine Internet causes an increase in the probability of having a bank in rural areas by 8.68 percentage points, versus 4.08 percentage points in urban areas.

In the context of sub-Saharan African countries, the number of financial institutions is significantly higher in urban areas than in rural areas. Indeed, economic activity being more important in urban areas, it is easy to understand that they have many more financial institutions than rural areas. This higher level of competition implies higher costs of entry into the financial market in urban locations. The need for financial products and services could therefore be fairly high in rural areas, so that a positive shock could generate more enthusiasm to offer such products and services in these areas where the relatively lower level of competition and the lesser costs at the entry of the market are likely to attract commercial banks.

2.6.3 Effects of Submarine Internet on Credit Access

We now investigate the effect of submarine Internet on the likelihood of obtaining loans from a financial institution. The estimates are shown in Table (2-7). Column (1) of Panel A shows that an individual is over 11 percent more likely to have obtained a bank credit after having access to submarine Internet. The positive effect of submarine Internet access on the likelihood of obtaining a bank credit is only positively significant in urban areas (Column (2)). The effect is a negative in rural areas, showing that submarine Internet did not benefit the average individual in these areas. In a later section, we show that, in rural areas, it only benefits business owners.

Figure (2-7) confirms the result that the arrival of the submarine Internet has caused an improvement in access to credit. Indeed, we can see that in the year preceding the treatment, the effect of the Internet on access to credit is statistically nil. However, in the post-treatment period, we have statistically positive effects.

The causal nature of Internet on access to credit is all the more apparent as we provide evidence that the estimated effects of submarine Internet are not governed by the distance range around the optical fiber network that we use to determine the connectivity status of different geographic positions. Figure (2-10) highlights estimates for various distances between the fiber optic network and the geographic location of potential loan applicants. We can see that access to finance is positive and increases as people move toward fiber optics, with significant effects up to 2 kilometers away. The people who remain unaffected by the Internet are those located more than 2.5 kilometers from the nearest optical fiber node (figure (2-10)). This fact not only supports the idea that our results are not sensitive to the range around the optical fibers, but also it supports the choice of 2 kilometers as the connection threshold. Indeed, figure (2-11) presents the coefficients of several regressions of the indicator of having received financing to start a business on different limits or thresholds. Beyond the threshold of 2 kilometers, the effect of Internet becomes negative. However, for connectivity defined at 2 kilometers or less, there is a statistically positive chance to have received bank credit to start a business (Figure (2-11)).

Effects of Submarine Internet on Businesses' Credit Access

In Panel B of Table (2-7), we replicate the same analysis as in the preceding section on the sub-sample of business owners. Firstly, we are interested in loan applications to finance business start-ups. We find that submarine Internet has larger effects on these individuals. The likelihood that an average business owner obtains a bank credit as a start-up capital increases by over 3.58 percentage points following access to submarine Internet (Column (1)). There is some heterogeneity in this effect by place of residence. Access to financing increases by around 6.10 percentage for business owners residing in urban areas and by 4.22 percentage points for business owners residing in rural areas following the arrival of submarine Internet. In other words, accessibility to submarine Internet positively affects access to financing, for both rural and urban entrepreneurs, with a higher and more significant effect among urban business owners. In addition, the fact that submarine Internet only positively affects business owners in rural areas means that the banks grant credit only to the owners of gainful activities in these areas.

Secondly, we assess the effect of submarine Internet on the probability of operational businesses receiving a loan approval. The results are presented in the table(2-8), where we see that an operational business, thanks to fast Internet access, has 4.77 percentage points more chance of obtaining financing approval compared to one that does not have access to high-speed Internet (column (2)). The magnitude of the effect of submarine Internet on the decision to grant credit is therefore 4.01 percentage points lower for business owners than for an individual in general. This result is consistent with the theory according to which, the possibility of offering a collateral increases the chances of the economic agent in need of financing to obtain a credit (Stiglitz & Weiss (1981), Benjamin (1978), Barro (1976)). Indeed, even though the company itself constitutes a collateral, it can be seen as a risky collateral, since any investment represent a risk; in comparison to the stable remuneration of a permanent worker which has a less risky characteristic.

The heterogeneity analyses in columns (3) and (4) of table (2-8) show that the effect of the internet on the propensity to obtain financing for owners of operational companies is 6.95 percent greater in urban areas compared to business owners not connected and rather lower by 8.86 percent in rural areas. The negative sign of loan approval to business owners in rural areas could be justified by the fact that it could be mainly agricultural credits. In fact, as in Hjort & Poulsen (2019b), we find that the arrival of high-speed Internet causes the migration of economic activity in rural areas from agriculture to sales and services occupations (table(2-12)). Precisely, we find that agricultural credit declines following access to submarine Internet. Assuming that, before the arrival of the submarine Internet, loans in rural areas were mainly agricultural credit, and considering that the arrival of the fast Internet has led to the migration of economic activities from the agricultural sector to other sectors of activity (Hjort & Poulsen (2019b)), then if the increase in loans in non-agricultural sectors is not proportional to the decline in agricultural credit, the net effect of the fast Internet on credit in rural areas will be negative. We discuss the effects on agricultural and non-agricultural credit in much more detail in the following subsection.

In order to validate the robustness of these results, we carry out a similar analysis on the data of businesses from the World Bank Enterprise Surveys (WBES). The data provide us with information on the origin of the financing, the movable and real estate assets of firms, as well as their ability of offering collateral as a guarantee for their loans, and much more on their access to information and to funding. We use these data to assess the impact of the advent of high-speed Internet on their propensity to obtain bank financing. Similarly to the analysis in Hjort & Poulsen (2019b), we consider an enterprise as connected to internet when the city in which it is located has at least one terrestrial fiber that passes through. Subsequently, accounting for the arrival dates of submarine internet in each country, we construct our variable "Fast Internet" by interacting the two variables. The regressions are therefore carried out on a sample of 11,807 companies based in 44 cities of 6 countries (Ghana, Kenya, Mauritania, Nigeria, Senegal and Tanzania).

The results are displayed in table (2-9). Columns (1) and (2) show that companies connected to high-speed Internet have 3.13 percentage points greater ability to offer collateral to guarantee their loans. This gives them approximately 6 percentage points better chances of obtaining bank financing (columns (3) and (4)), in comparison to non-connected businesses. This estimate is in fact consistent and very close to that (4.77 percentage points) obtained from the LSMS sample table (2-8).

In addition, the WBES collects information on evaluation of access to finance. Using this information, we find in columns 5 and 6 that, compared to unconnected businesses, enterprises with high-speed Internet access rate access to finance to be 5.2 to 5.82 percent better. Lastly, access to fast Internet has a positive effect on the percentage of capital obtained from banks, with around 2 percent of their capital owned by formal commercial banks.

2.6.4 Effects of Submarine Internet on Access to Agricultural versus Non-agricultural Credit

In this subsection, we first analyze how access to the fast internet affects individuals' perceptions of accessibility to agricultural and non-agricultural credit. Second, we examine the effect of fast Internet on access to credit, based on the actual financing approval decisions.

Effects of Submarine Internet on Access to Agricultural versus Non-agricultural Credit: Individuals perceptions

In the LSMS surveys, individuals are asked to rate progress in accessibility to agricultural and non-agricultural credit, as well as access to financial advice. For each of these three elements, we define a variable that takes the value of one when the individual's assessment is in favor of improved accessibility and zero when the opinion is rather unfavorable. We analyze the effect of submarine Internet on these outcome variables. We estimate equation (2.1), now replacing the dependent variable by each of these three outcome variables. The results are reported in Table (2-14)).

In Column (1) of Panel A, we find that the likelihood that an average person reports an improvement in the accessibility to agricultural credit increases by 4.13 percentage points after the arrival of submarine Internet. Column (2) shows that this effect is large in urban areas (2.64 percentage points). In rural areas (Column (3)), access to submarine Internet decreases the likelihood of reporting improvement in the accessibility of agricultural credit by 13.1 percentage points. This negative effect is somewhat puzzling, given that agricultural activities constitute the main income-generating activities of rural areas in Africa. A possible explanation is that rural residents shift to non-agricultural activities following the introduction of high-speed Internet, which would be consistent with the negative effect of submarine Internet on low-skill employment and its positive effect on high-skill employment documented in Hjort & Poulsen (2019b). A decreased involvement in agricultural activities following the introduction of submarine Internet would then decrease the demand for agricultural credit, explaining our findings.

In Panel B, we analyze the effect of submarine Internet on the evaluation of the access to non-agricultural credit, and find a positive effect. Submarine Internet increases the likelihood of reporting improvement in access to non-agricultural credit by 7.77 percentage points in the whole sample. The effect is greater in rural areas (+15 percentage points) versus urban areas (+2.29 percentage points). The fact that the effect is positive in rural areas contrasts with the negative effect on agricultural credit documented in Panel A, lending credence to the fact that submarine Internet causes a shift away from agricultural activities.

In Panel C, we analyze the effect of submarine Internet on access to advice about accessing agricultural credit. The results largely reflect those found in Panel A. Submarine Internet increases the likelihood of reporting improved quality of financial advice in the whole sample by 0.30 percentage points. However, this positive effect is only present in urban areas (+2.56 percentage points), being negative in rural areas (-2.64 percentage points). The findings confirm that submarine Internet has only positively impacted the financing of non-agricultural activities in rural areas.

Effects of Submarine Internet on Access to Agricultural versus Non-agricultural Credit

Fast Internet significantly deepens development in rural areas through the move of the labour force from the agricultural sector to other sectors of activity. We can appreciate this at at least two levels. First in table (2-6), we show that rural communities connected to high-speed Internet are 8.68 percent more likely to have a banking institution in their precincts, compared to 4.08 in urban communities. Secondly, from the results of table (2-10), we see that, in comparison to rural areas[10], the effect of the fast Internet on access to credit for non-agricultural business owners is greater in rural areas than in urban areas; non-agricultural business owners in rural areas who are connected to high-speed internet are 11.2 percent more likely to get a loan approval, while those in urban areas are only 3.47 percent more likely to receive financing. The effects on agricultural credit are significantly lower.

Our results are consistent with those of Hjort & Poulsen (2019b), which show that fast

[10] where employed workers migrate from the agricultural sector to the sales and trading sector; or existing agricultural businesses turn to sales, but also new businesses are created due to the advent of the fast Internet are also sale oriented.

Internet allows the migration of labor from the agricultural sector to the non-agricultural sector in rural areas (table (2-13)). We can therefore understand why not only the effect on non-agricultural credit is found to have a greater magnitude in connected rural areas compared to connected urban areas, but also why agricultural credit has been reduced since the arrival of submarine internet (table(2-10)). The negative sign in column (4) of table(2-8) would simply reflect the fact that, the weight of firms that are involved in the decrease of the agricultural credit is greater than that of firms involved in the increase in non-agricultural credit in rural areas.

At last, it is interesting to note that agricultural credit in urban areas are 6.19 percentage points higher following connection to high-speed Internet. Note that agriculture is not the main activity of urban areas. Agricultural companies in urban areas may mainly be companies that export agricultural products, and therefore have a better rating, than agricultural businesses in rural areas, and thereby better creditworthiness. In fact, the access to high-speed Internet could give them better visibility, allowing them to prospect and expand their market. The resulting increase in income could therefore constitute an additional guarantee to the loan, favoring approval of their loan applications.

If banks grant more loans under the influence of submarine Internet, then it is evident that sub-Saharan African financial institutions are less reluctant to finance businesses since the introduction of this technology. In the next section, we confirm this finding by analyzing how bank liquidity has responded to the advent of submarine Internet in Africa.

2.6.5 Submarine Internet and sub-Saharan African Banks' Excess Liquidity

The graphs in Figure (2-12) give us a clear picture of the relation between connectivity to the Internet and the financing of the economy (at the country level). Both the volume of credit granted to the private sector and the number of internet users have grown significantly in the last two decades, although the growth speeds are different.

Likewise, bank excess liquidity decreases with the number of secure internet servers. Here also, the growth rate of the internet is clearly superior to that of excess bank liquidity in absolute terms. In addition, the slope of the domestic credit curve is much steeper than that of bank excess liquidity. This suggests that the arrival of submarine internet reduces the reluctance of banks to grant loans, but the amounts of loans granted are not increased enough to significantly reduce bank excess liquidity. Is this just a correlation, or could this relationship be interpreted as causal?

We use a World Bank's dataset on the number of secure internet servers existing in each country for 32 sub-Saharan African countries to examine this question. A country is connected to the Internet when it has at least one internet secure server. In addition, using the date of arrival of submarine internet in each country, we construct a variable which takes the value of one from the date of arrival of submarine internet in that country, and the value zero in all the preceding periods. We thus calculate the differential between the impact of internet on bank excess liquidity before and after the arrival of fast internet by distinguishing the sample that did not have access to fast internet from that which is connected to this technology. In summary, a country is connected to submarine Internet if it has at least one secure internet server and if its Internet backbone network is connected to the terrestrial fiber nodes. We therefore use the variable that indicates whether the country is connected to fast internet to determine its effect on the reluctance of banks to grant loans and also on bank excess liquidity through credit granted to the private sector by estimating the following equation:

$$CRT.ELIQ_{ct} = \beta_0 + \beta_1 Submarine_{Ct} + \beta_2 GDP_{ct} + \sigma_c + \gamma_t + \epsilon_{ct} \tag{2.2}$$

Where $CRT.ELIQ_{ct}$ is the domestic credit to the private sector (resp. bank excess liquidity); $Submarine_{Ct}$ is the connectivity of country c at time t to fast internet; and GDP_{ct} is the gross domestic product (GDP) of country c at time t. We also control for country fixed effect (σ_c) and year fixed effect (γ_t).

The results of the estimation of equation(2.2) are represented in table (2-15). We find that following connection to submarine Internet, the liquidity-to-assets ratio of a country significantly. In other words, access to high-speed internet causes the decline in bank excess liquidity as a ration of their total assets by 4.24.

More interestingly, we also investigate the impact of high-speed Internet on the volume of credit granted to the private sector, and we find that a country connected to high-speed Internet has about 4.9 percent more credit granted to the private sector compared to an unconnected country, in the same period.

In summary, access to submarine Internet increases the propensity of banks in sub-Saharan Africa to lend to the private sector. As noted above, there is a difference in magnitude between the effect of fast Internet on credit to the private sector and its effect on bank excess liquidity. This difference could be explained by the fact that many factors justify banks' reluctance to grant loans; however, fast Internet access eliminates some of these factors. In the next section, we document these factors.

2.7 Possible Mechanisms

We identify four possible channels through which fast internet decreases banks' reluctance to grant loans. We argue that, by increasing access to information, enhancing individual and firm productivity, improving the solvency of potential borrowers, and by favoring employment and entrepreneurship, access to fast internet causes financial institutions to be less reluctant to grant credit. We explore each of these channels below.

2.7.1 Access to Information

In this section, we show how the connectivity to submarine internet facilitates banks' accessibility to information on their potential borrowing clients. Indeed, we show that submarine internet access allows banks, in sub-Saharan Africa to obtain information more efficiently about the solvency of their clients, which is information that would allow them to take the most informed decision towards their clients' demands for financing.

In figure (2-13), on the left, we can see that before the arrival of fast internet, the average google search for the expression "credit score" was constant at around 2.[11]. After the arrival of the fast internet, it presents an upward trend with a steady growth rate. This suggests a causal relationship between the advent of submarine internet and the increased interest in credit records. This presumption of causation can be confirmed by the figure on the right, where we compare South Africa to Canada.

Despite the fact that there is a slight negative slope on the curve of Canada between 2004 and 2007[12], it should be noted that over the entire period from 2005 to 2020, Canada was connected to high-speed internet. By contrast, South Africa only acquired connectivity to the submarine internet in the third quarter of 2009. It is interesting to note that the gap between Canada and South Africa regarding the search for the expression "credit score" considerably reduces after the arrival of fast internet, converging towards zero. This is consistent with the hypothesis of a causal effect of the arrival of submarine internet on the growing search for information on financial solvency. Thus, fast Internet, by facilitating access to information

[11]The numbers on the Y axis correspond to search interest relative to the maximal value on the chart for a state at a specific time. A number equal to 100 indicates the highest popularity for the term. A rate of 50 indicates that the word search for that period is half as trendy. A record of 0 implies that the data is missing for that month.

[12]This negative slope could be explained by the fact that this period coincides with the economic recession that North American economies experienced, following the explosion of the dot-com bubble in the early 2000s. Known as the Internet bubble, it was defined as a stock market bubble generated by an extreme speculation of Internet-related companies at the end of the 90s. This period was characterized by substantial growth of interest in the usage and acceptance of the Internet. Technology companies filed for bankruptcy afterwards. This, in fact, could have shifted the popularity of the term "credit score" in favor of that of the term "dot-com" on Google, which precisely recorded 100% popularity in 2004 in Canada.

on the reliability of borrowers through their credit report, would influence the propensity of banks to finance economic agents in need of financing, and thereby affects their excess liquidity.

The micro-dataset of businesses of the World Bank (WBES) also inform us about the place of the Internet in the accessibility of firms to information necessary for the smooth running of their operations and production activities. In particular, it tells us if a given company has a website, if telecommunication is an obstacle to the smooth running of their operations, what are the sources of information that have enabled them to find their new supplier and also the level of complexity of credit application procedures. We use these four indicators to assess the impact of Internet access on the availability of information critical to the credit and financing market.

The results of the estimations of equation(2.1) on this sample are reported in table (2-16). In the first column, a company connected to submarine Internet is 3.83 percentage points more likely to have a webpage compared to one that is not connected. Indeed, a company's website, by allowing it to increase its marketing possibilities at a lower cost, but also by increasing its visibility, gives it better credibility than a company not connected to the Internet, and therefore gives it a better chance of obtaining a positive decision to its funding application. Moreover, there is an obvious relation between this result and that of column (2), showing that firms with Internet access are 6.95 percentage points less likely to find telecommunication as an obstacle to their operations. Column (3) also shows how the Internet facilitates access to information critical to market efficiency, conferring 7 percentage points greater chance of obtaining information on potential suppliers for connected firms. Likewise, it allows better access to information useful for preparing a credit application. By facilitating access to this type of information, the Internet reduces the complexity of the credit application process by 3.29 percentage points (column (4)).

In summary, the Internet, by facilitating access to information, reduces inefficiencies in the banking markets of sub-Saharan Africa, thereby reducing the reluctance of commercial banks to finance the private sector.

2.7.2 Submarine Internet and Productivity

We argue that submarine Internet increases the likelihood of receiving a loan by increasing productivity. The existing literature shows that the productivity of companies facilitates their accessibility to financing in the financial markets (Olawale & Akinwumi (2010), di Mauro et al. (2017)). Another recent body of works demonstrates that Internet connectivity improves productivity and job creation (Espinoza et al. (2020), Hjort & Poulsen (2019b), Rollo & Paunov (2015), Wagner & Cockburn (2010), Batjargal (2007), Cockburn & Wagner

(2007)). These two bodies of literature, put together, imply that the advent of fast Internet, by improving individual and firm productivity as well as employment, would increase income and make it easier to acquire assets (real estate, materials and equipment, etc.) that can be used as collateral for a loan. This in turn would reduce the reluctance of banks to grant loans. Indeed, the literature on the excess liquidity of African banks in sub-Saharan Africa argues that a lack of guarantee or collateral is one of the reasons why banks in this region are very reluctant to grant loans (Elali et al. (2013), Kira & He (2012), Barbosa & Moraes (2003), Coco (2000)).

The World Bank's enterprise data allow us to validate this hypothesis. Table (2-17) therefore includes the results of these estimates. In the first column, connected firms have national sales revenue as a percentage of annual sales 3.39 points higher than that of unconnected firms. However, direct exports, which represent only 3.79 percent of annual sales, although having a positive coefficient of 0.45 percentage points, are statistically similar across connected and unconnected firms (column (2)).

We also find that the volume of losses as a percentage of annual sales is 5.27 points lower for connected firms compared to unconnected ones (column (3)). In the 4th column, the productivity of the firms is obtained by computing the ratio of the annual sales to the expenses of remuneration of the labor force. We find that Internet access increases productivity by 19.2 percentage points.

Thus, productivity is a factor that favors the reduction of banks' reluctance to finance the private sector. Indeed, by increasing the productivity of firms, it gives them better prospects to secure their loan with a collateral.

2.7.3 Submarine Internet, Employment and Entrepreneurship

We also argue that submarine Internet increases the likelihood of receiving a loan by increasing employment opportunities and entrepreneurship. We test this hypothesis by analyzing the causal effect of submarine Internet on entrepreneurship and employment in sub-Saharan Africa. Results are reported in table (2-18). Column (1) of Panel A shows that the likelihood of securing a permanent employment position increases by 4.15 percentage points following the arrival of submarine Internet. This result is consistent with the findings of Hjort & Poulsen (2019b), who found that the arrival of fast Internet in sub-Saharan Africa increases an individual's chance to be employed by a factor ranging from 2.7 to 7.7 percentage points, depending on the sample used to conduct the analysis.[13] Moreover, we find that this positive effect is also present in urban settings, where submarine Internet increases the likelihood of

[13] Hjort & Poulsen (2019b) use Demographic and Health Surveys, Afrobarometer surveys, and the South Africa Quarterly Labour Force Survey.

a permanent employment position by 2.71 percentage points (Column (2)). The effect is higher in rural areas (13.1 percentage points in Column (3)).

In Panel B, we analyze the effect of submarine Internet on the number of hours worked per week, finding a large and significant positive effect. We find that the number of hours worked per week increases on average by 1 hour and 53 minutes, following the arrival of submarine Internet (Column (1)). Again, this effect is very heterogeneous among urban and rural residents. In urban areas, an individual connected to submarine Internet has approximately 4 more hours of work per week (Column (2)), whereas in rural areas, it has only 0.30 more hours of work per week (Column (3)). This again confirms that access to this communication and information technology does not help agricultural activities, which are the main activities in rural Africa.

In Panel C, we test the hypothesis that submarine Internet promotes entrepreneurship in sub-Saharan Africa. We estimate that an individual is over 2.65 percentage points more likely to run their own business following access to submarine Internet (Column (1)), implying that this technology favors entrepreneurship. There is some heterogeneity in the effect of submarine internet on entrepreneurial incentives by place of residence. Access to submarine Internet increases the likelihood of running ones own business by 1.79 percentage points in urban areas and versus 4.1 percentage points in rural areas. Therefore, while the effect is larger in rural areas, the difference is quite interesting. When compared to the findings of Panels A and B, this latter finding suggests that submarine Internet induces differential processes of job creation in urban and rural locations. While additional jobs are created in urban areas in response to exposure to fast Internet, in rural areas, this technology not only causes a shift away from agricultural activities into sectors that are more skilled, but it also create more permanent employment.

Finally, we establish directly that having a stable job facilitates access to credit. To do this, in our baseline model (equation 2.1), we replace the dependent variable by a binary variable indicating whether an individual has obtained a bank loan or not, and the independent variable by a binary variable indicating whether an individual has a permanent job. The results are reported in Panel A of Table (2-19). We find that an individual who has a permanent job is 88.4 percentage points more likely to obtain a bank loan, compared to one who does not have a permanent job (Column (1)). This effect is larger in rural areas (89.7percent) than in urban areas (85.9 percent). The effect for rural areas implies that it is only the nature of activities carried out in these areas that explains the negative effect of submarine Internet on access to bank credit found in these areas (Column (3) of Panel A in Table (2-7)).

In panel B, restricting the analysis to the sample of firms' owners, we find higher effects in urban areas. Business owners having a permanent position in urban areas are 50.8 percent more likely to have obtain a bank credit to launch their businesses, versus 43.0 percent for rural business owners.

All together, the findings show that fast internet access expands the probability of obtaining bank financing partly through the positive and greatly significant impact it has on employment opportunities and entrepreneurship. This technology gives permanent workers and entrepreneurs a better prospect of obtaining a collateral guaranteeing their bank financing.

2.7.4 Submarine Internet and Solvency

Another way through which access to submarine Internet would decrease the reluctance of financial institutions to provide funding is by increasing the solvency of borrowers. Indeed, the credibility of a borrower increases his chances of obtaining a favorable response to a loan application. In other words, an entrepreneur with a good credit rating is more likely to get a future loan, compared to one who has a relatively low credit rating. We argue that, by increasing entrepreneurs' productivity and income, fast Internet access increases their ability to repay their loans, therefore increasing their credibility and their future borrowing potential. We test this hypothesis and report the results in table (2-20).

We find that access to fast Internet increases the probability that a borrower will repay their loan by over 16 percentage points (Column (1)). In addition, the way in which access to submarine Internet impacts the creditworthiness of borrowers varies, depending on whether one lives in a rural or an urban area. We find that fast Internet positively and very significantly promotes the likelihood of repaying a bank loan by 4.41 percent in urban areas versus 1 percentage points in rural areas. The larger effect in urban areas reflects the relatively stable nature of employment in comparison to seasonal employment in rural areas.

2.8 Conclusion

Small and medium-sized enterprises account for over ninety five percent of businesses and over sixty percent of employment in the private sector in sub-Saharan African countries. Yet these businesses have great difficulty in obtaining loans from banks, despite the fact that the latter generally have significant excess liquidity. This chapter examines how the advent of fast Internet has affected credit access in this region, and suggests implications for commercial bank's excess liquidity. We combine geo-spatial information on Internet with individual-level data from the Living Standard Measurement Study and other datasets for the analysis.

Exploiting the gradual arrival of submarine Internet cables, we use a difference-in-differences framework and find that this technology significantly increases credit access in sub-Saharan Africa both for individuals and firms. Using firm-level data, we show directly that fast Internet increases the probability of a firm acquiring assets that can be used as a collateral. Using country panel data, we show that banks' willingness to grant loans has increased, and their excess liquidity has significantly decreased in response to the introduction of fast Internet.

We find significant regional heterogeneity in the effect of submarine Internet on loan access. This effect tends to be larger in urban areas. In rural areas, access to submarine Internet decreases loan access for the average individual, but we find a positive effect on entrepreneurs. Our analysis further distinguishes between agricultural and non-agricultural loans. We find large positive effects, but the effect on agricultural loans is negative in rural areas, which is intriguing in light of the fact that rural economies are primarily agricultural. The findings suggest that fast Internet primarily promotes the expansion of the non-agricultural sector, and a shift away from agricultural activities.

We document possible mechanisms through which submarine Internet affects credit access. Based on data from Google Trends, our findings suggest that this technology facilitates the online acquisition of information of the creditworthiness of potential borrowers. In addition, using our main data source, which is the LSMS, we find that fast Internet promotes both entrepreneurship and permanent employment and enhances productivity, thereby increasing income and the ability of individuals and firms to repay loans. However, the effects on permanent employment and productivity are only positive in urban areas. We argue that while additional jobs are created in urban areas in response to exposure to fast Internet, in rural areas, this technology only reallocates labor away from agricultural activities and into sectors that are more skilled, resulting in a net loss in permanent employment. Urban economies are more stable by nature. This stability is enhanced through the positive impacts of submarine Internet, which in turn increases the solvency and creditworthiness of urban residents. Indeed, we find that submarine Internet increases the likelihood of repaying a loan, and that this effect is larger in urban areas.

Our analysis suggests that the expansion of communication technologies in Africa should be combined with policies to increase not only financial market's efficiency, but also to guarantee some benefits to the whole economy.

2.9 Appendix

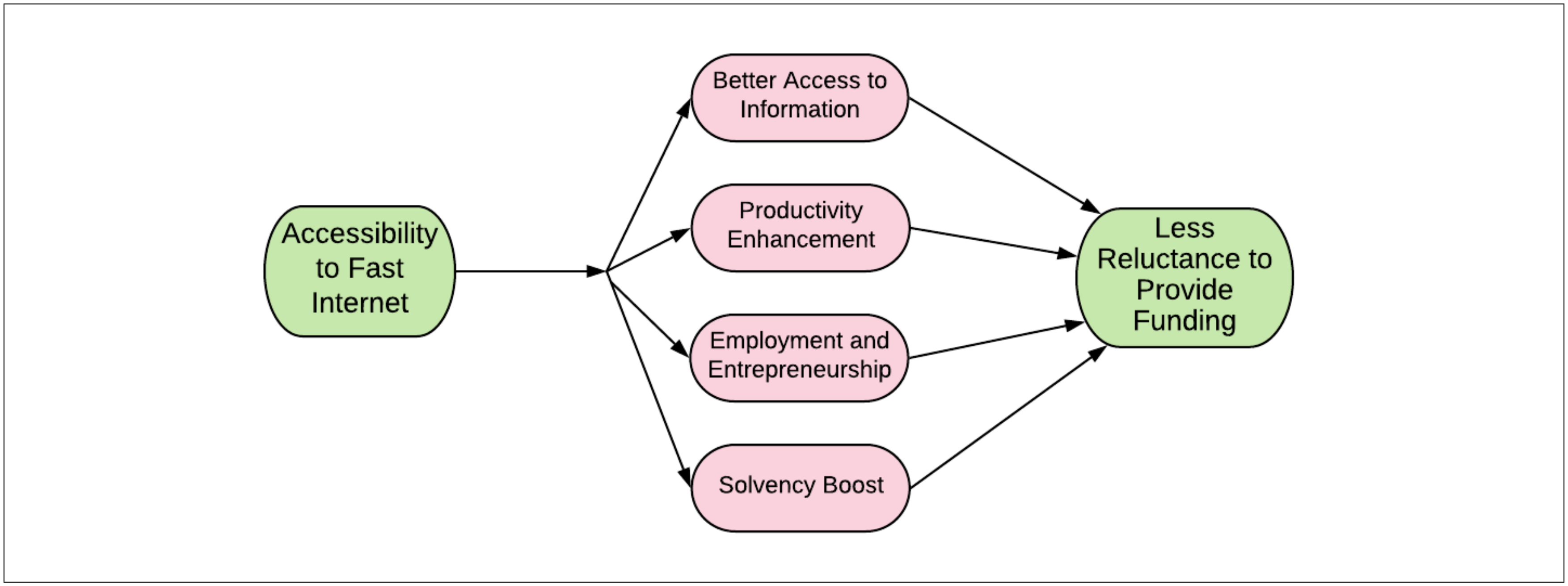


Figure 2-1: Conceptual Framework.

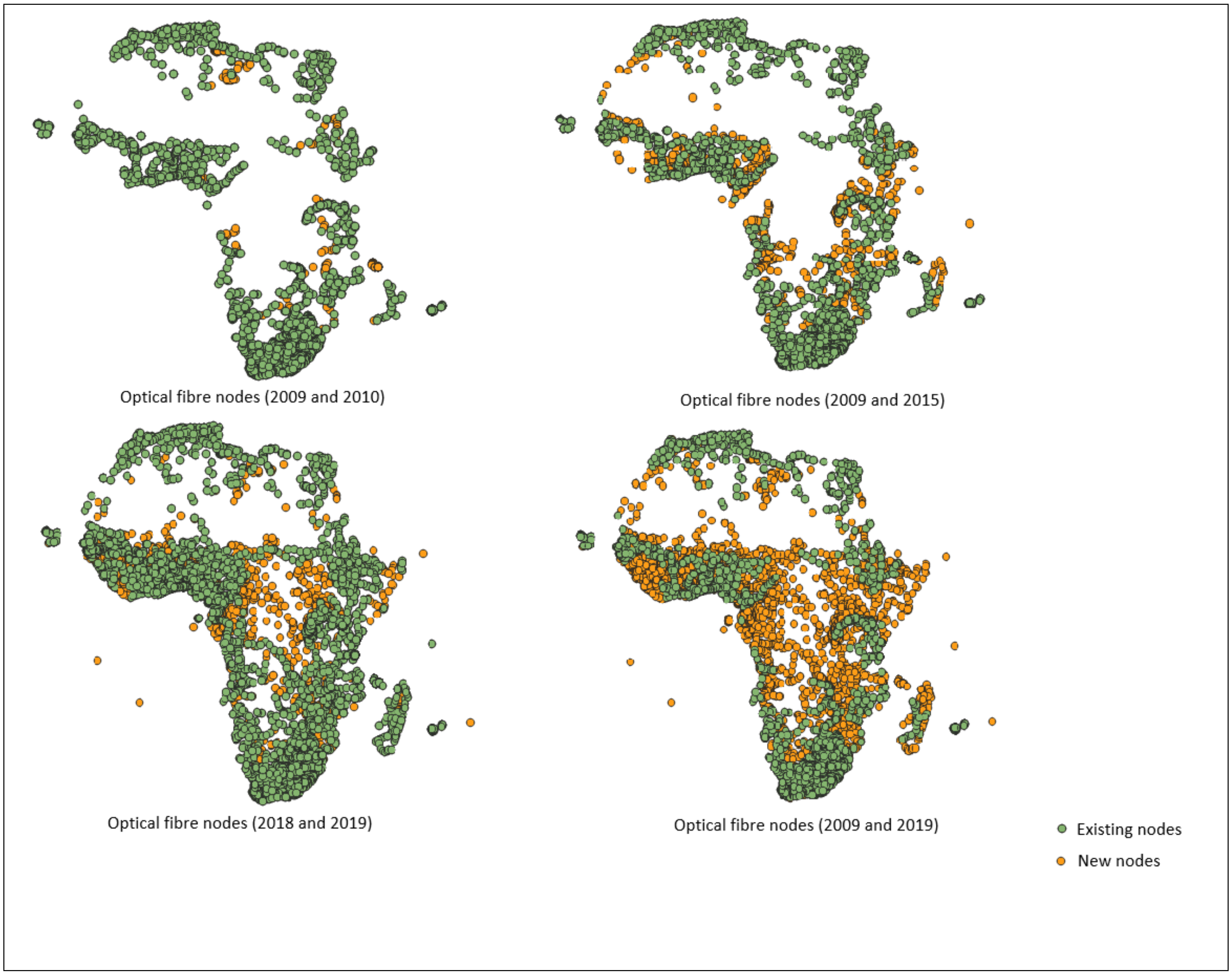


Figure 2-2: Coverage of optical fiber nodes between 2009 and 2019.

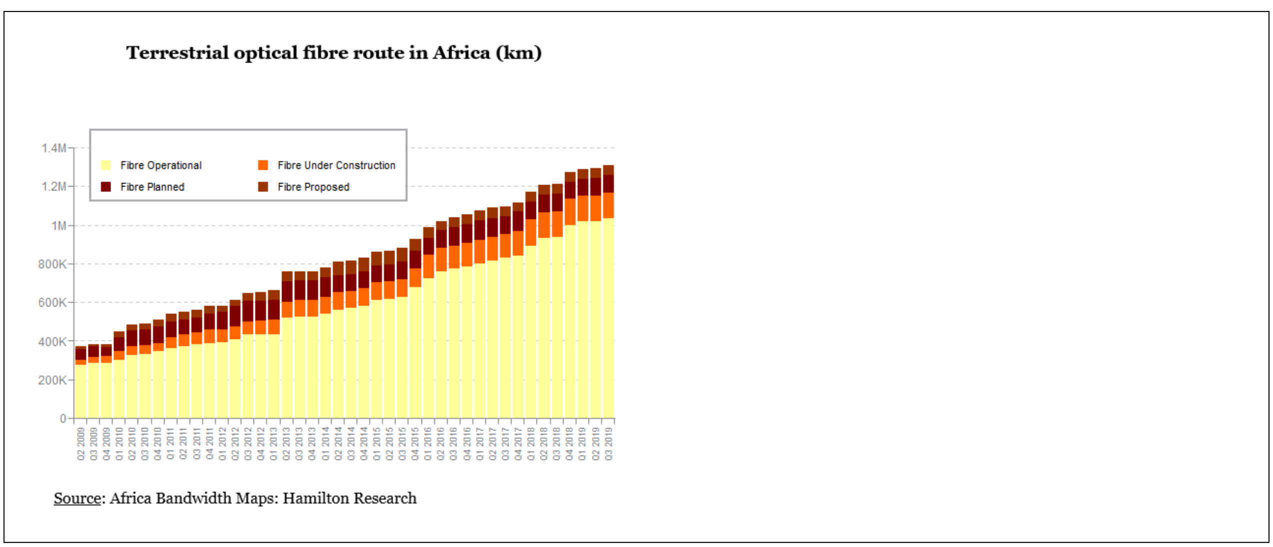
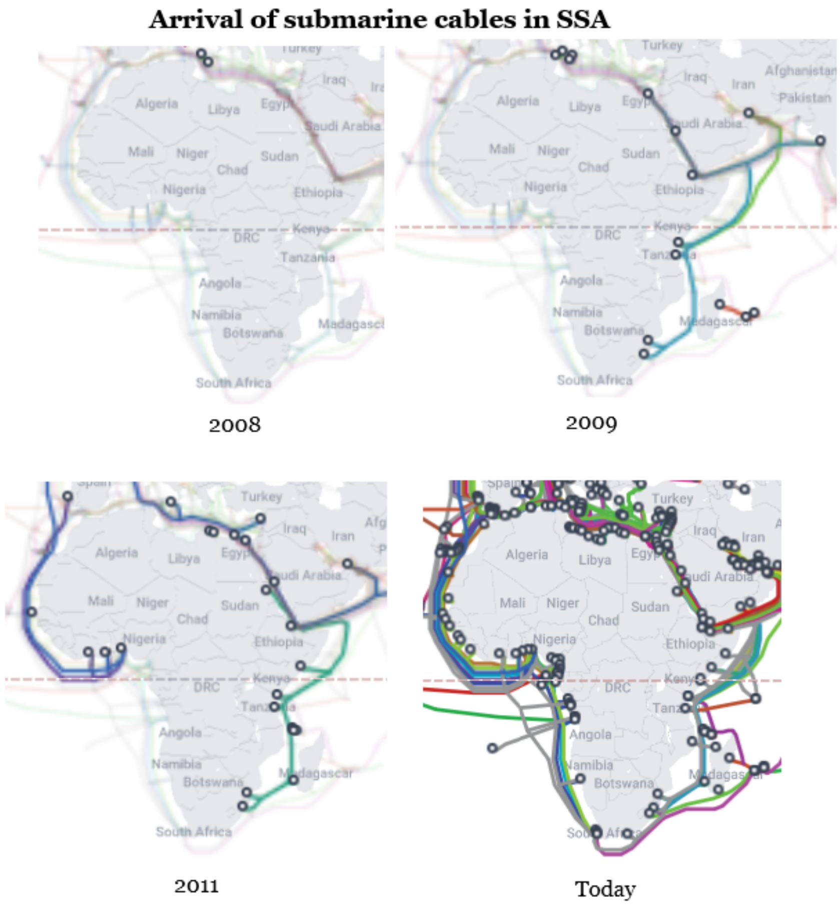


Figure 2-3: Fiber Optical Fiber Network Expansion.



Source: TeleGeography

Figure 2-4: Gradual Arrival of Submarine Cables in Africa.

	Type	Core / Cladding (um)	Fast Ethernet 100Mb	Gigabit GbE	10Gigabit 10GbE	40Gigabit 40GbE	100Gigabit 100GbE
Multimode	OM1	62.5 / 125	2km	275m	33m	-	-
	OM2	50 / 125	2km	550m	82m	-	-
	OM3	50 / 125	2km	800m	300m	100m	100m
	OM4	50 / 125	2km	1100m	400m	150m	150m
Singlemode	OS1/OS2	9 / 125	40km	100km	40km	40km	40km

OM: Optical Multimode

OS: Optical Single Mode

Source: Teleweaver Technologies

Figure 2-5: Fiber Optical Fiber Emission Distances.

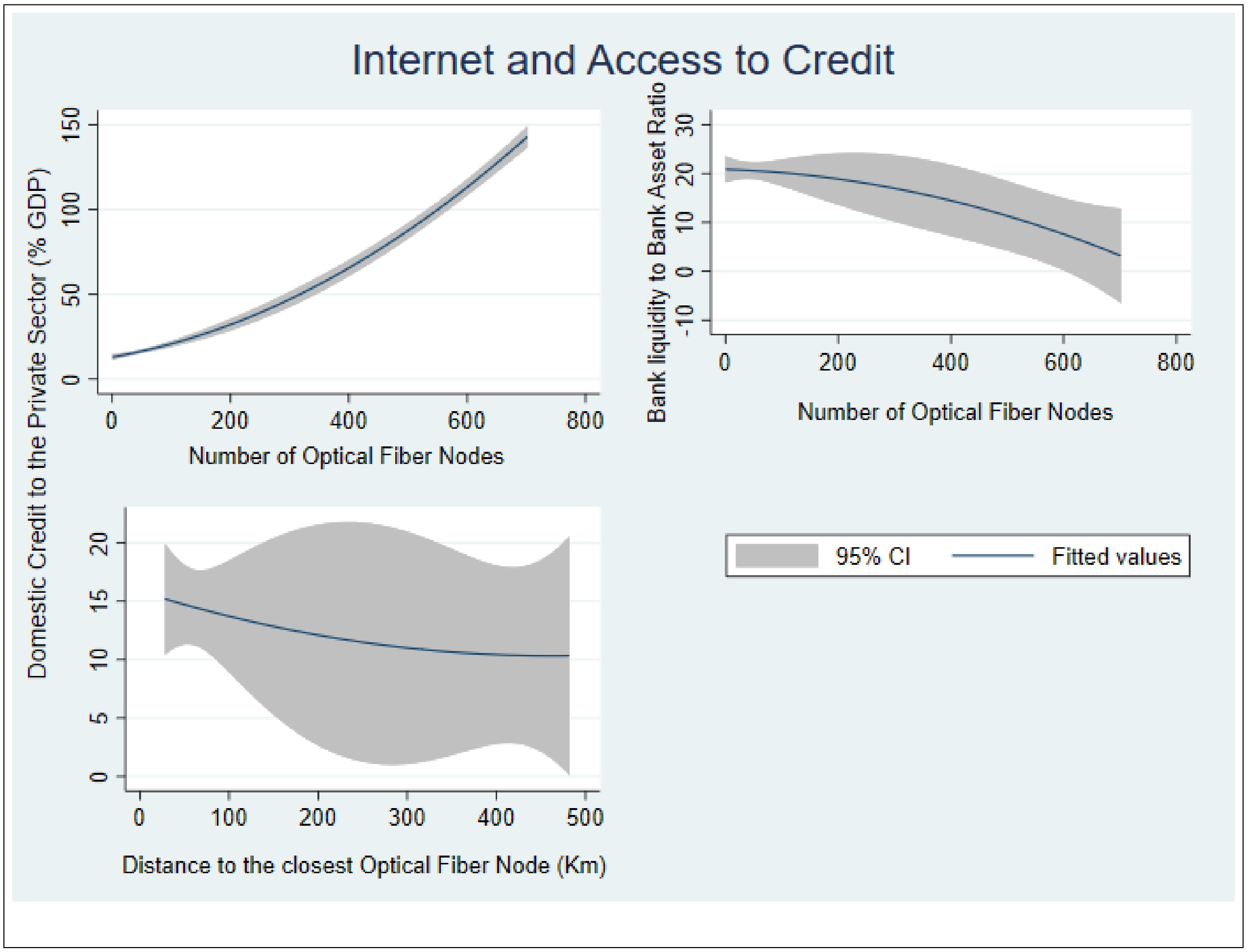


Figure 2-6: Internet and Access to Credit.

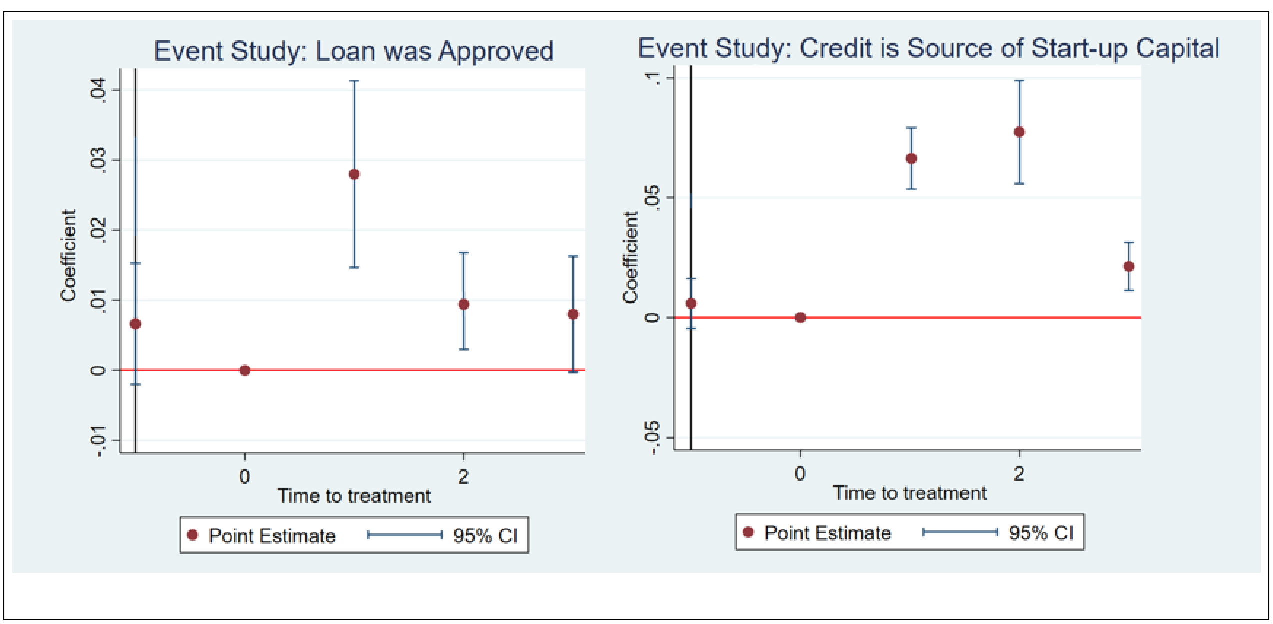


Figure 2-7: Event study (LSMS).

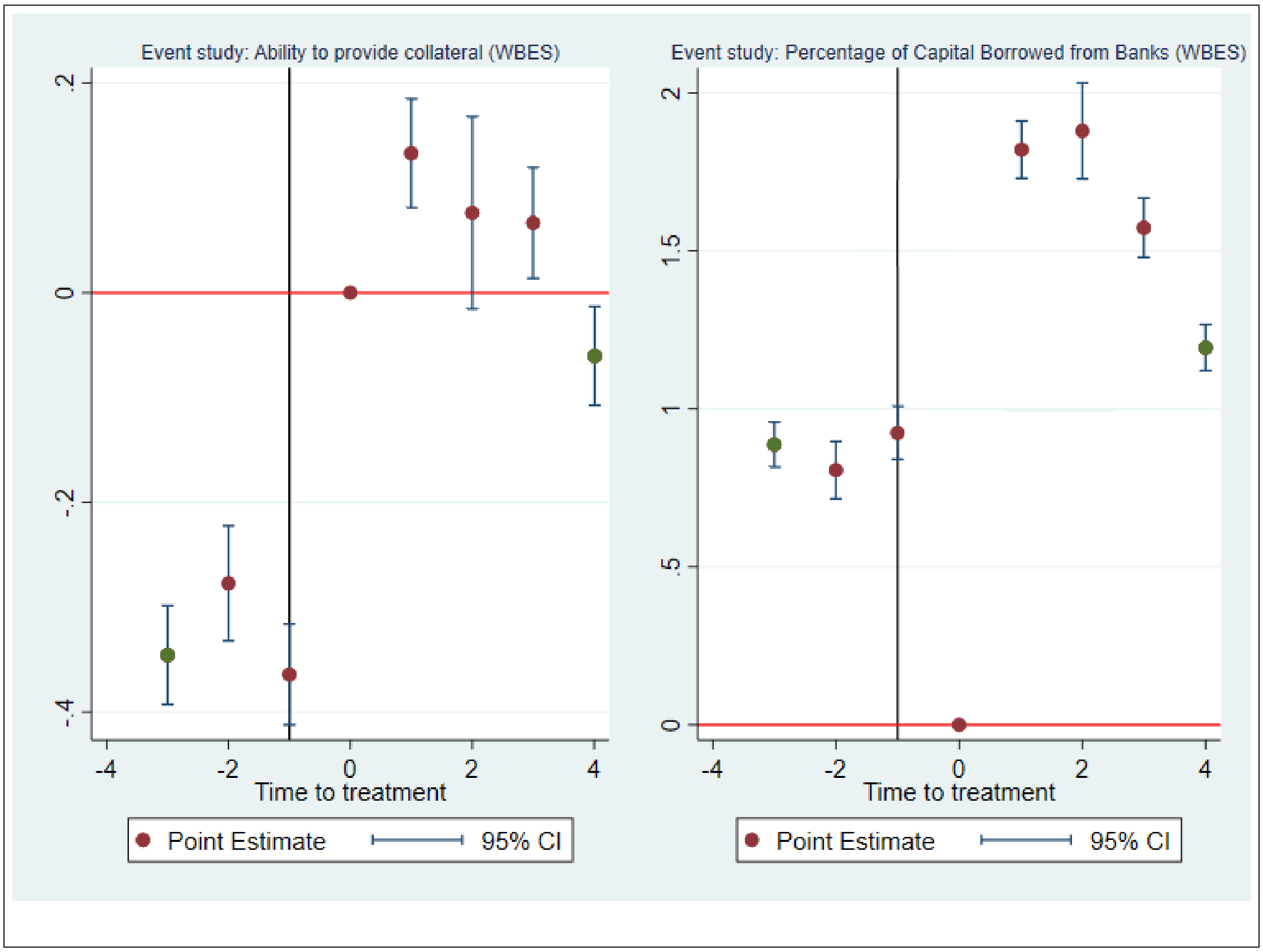


Figure 2-8: Event study (WBES).

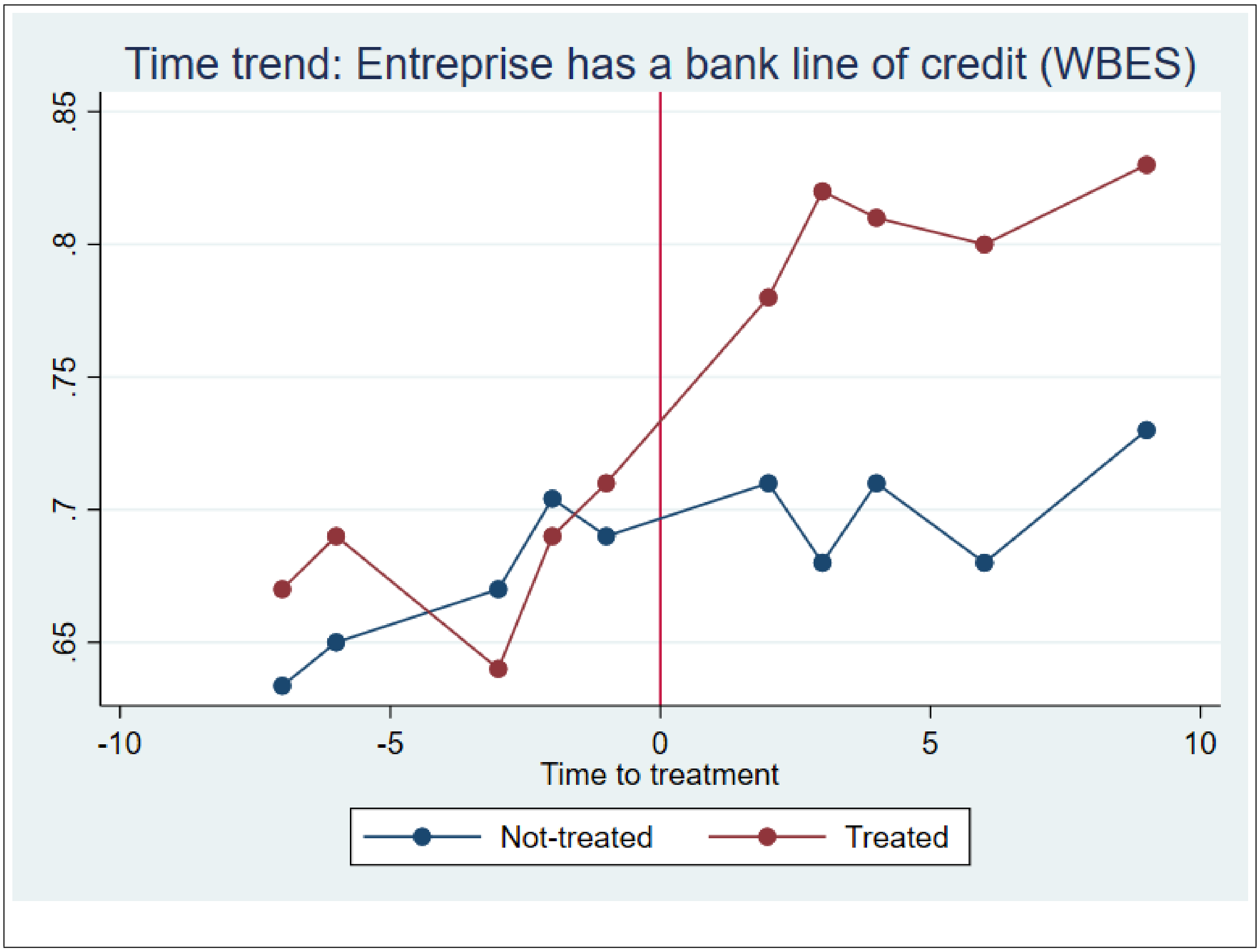


Figure 2-9: Time Trend in Enterprises Likelihood to have a Bank Line of Credit (WBES).

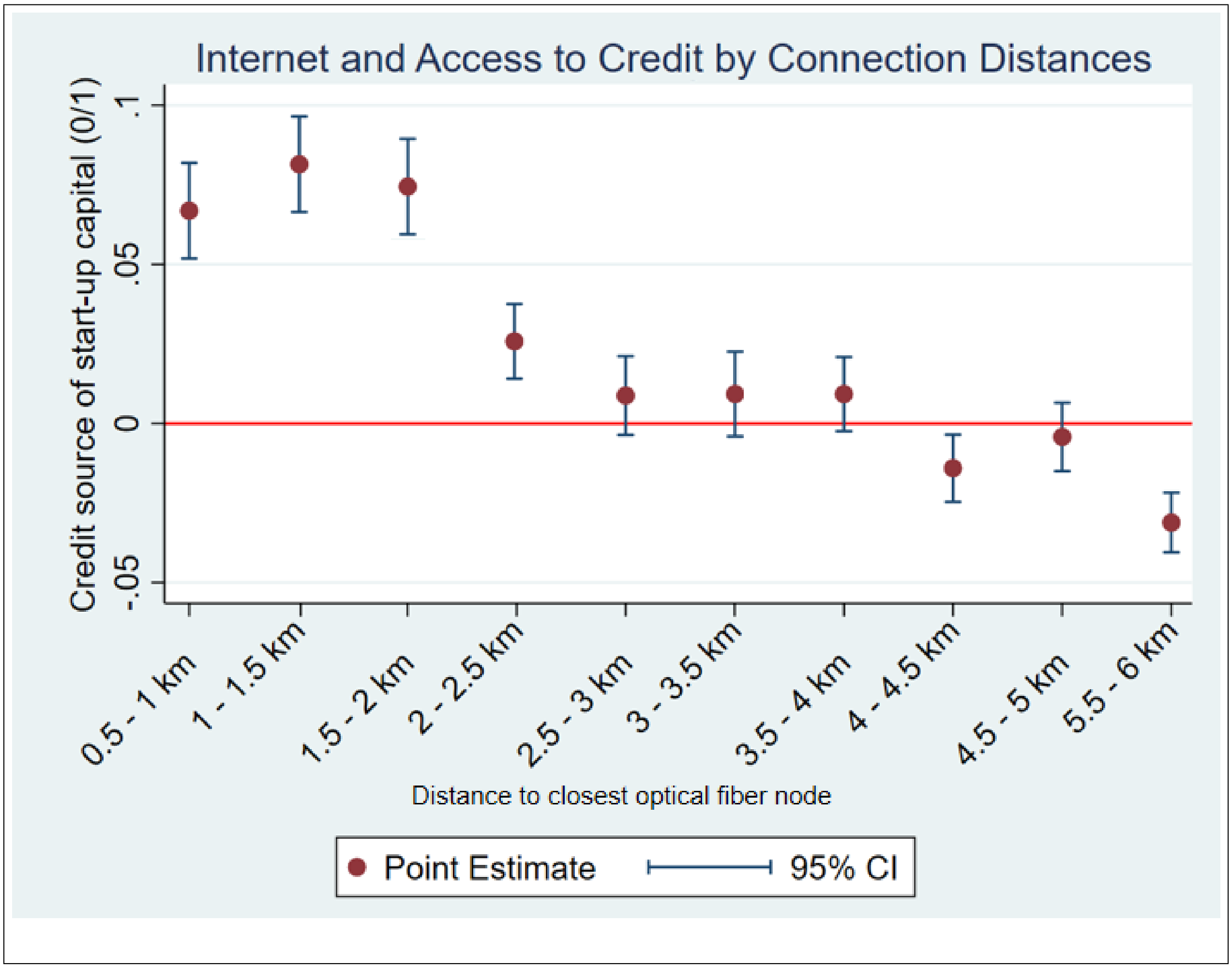


Figure 2-10: Internet and Access to Credit.

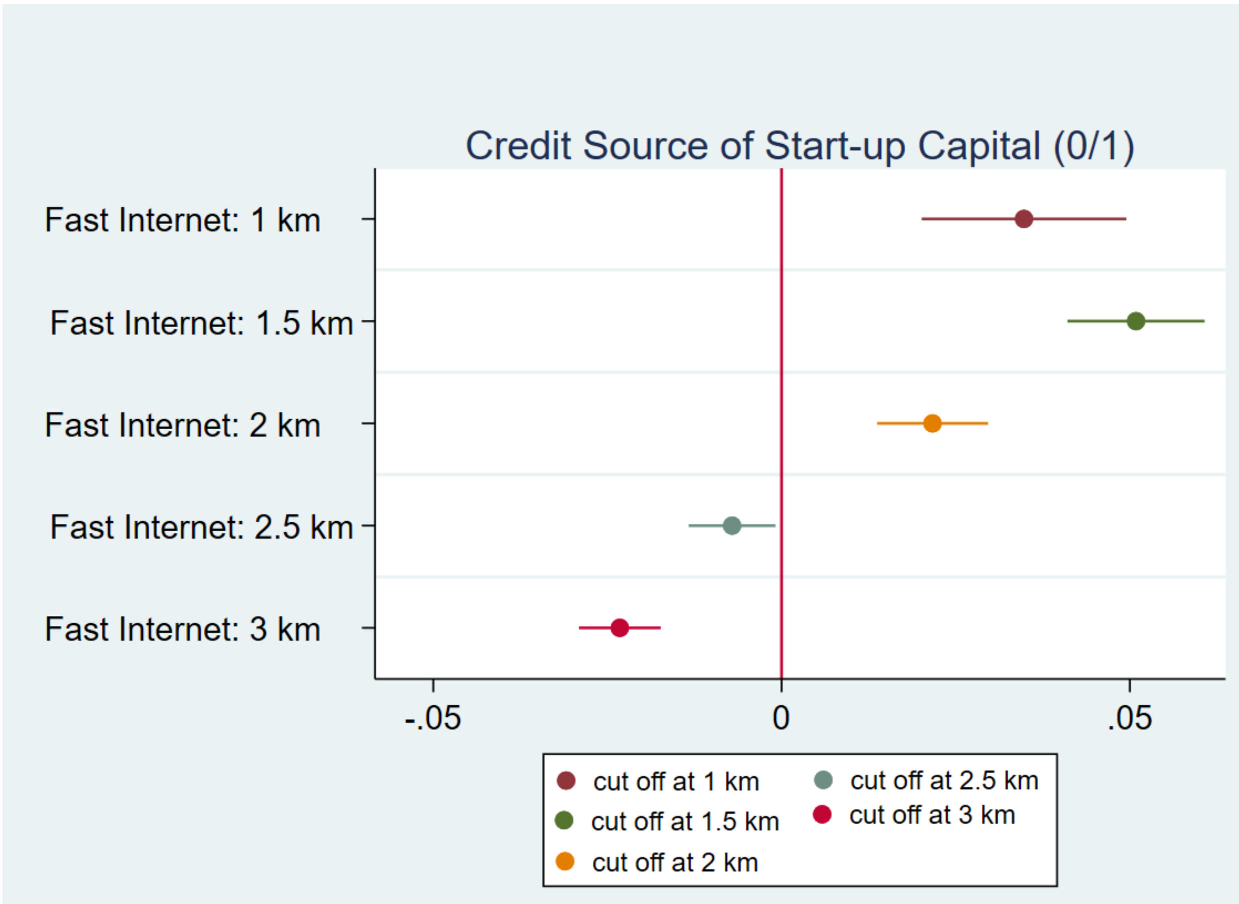


Figure 2-11: Internet and Access to Credit for Different Cut-Off.

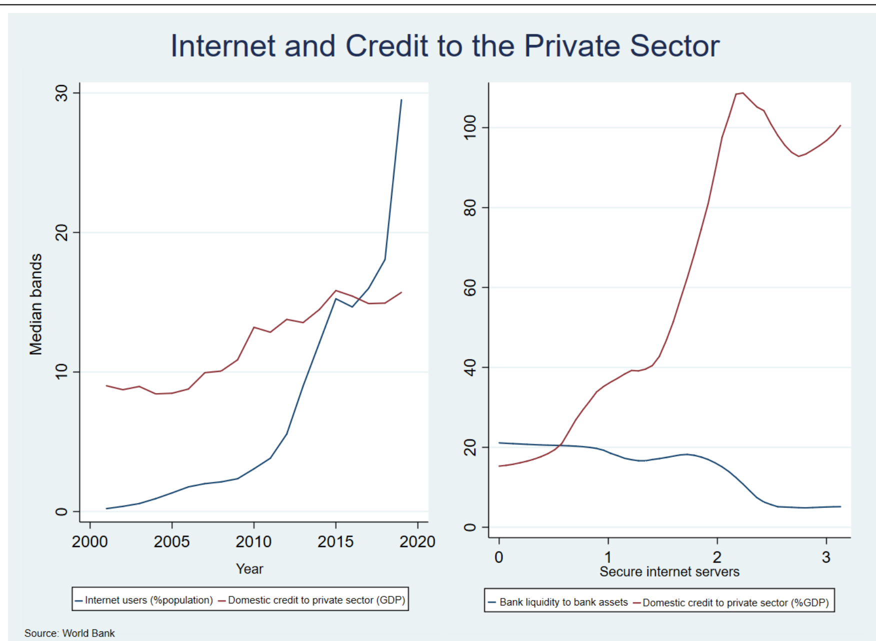


Figure 2-12: Access to Internet, Credit to the Private Sector and Excess Liquidity.

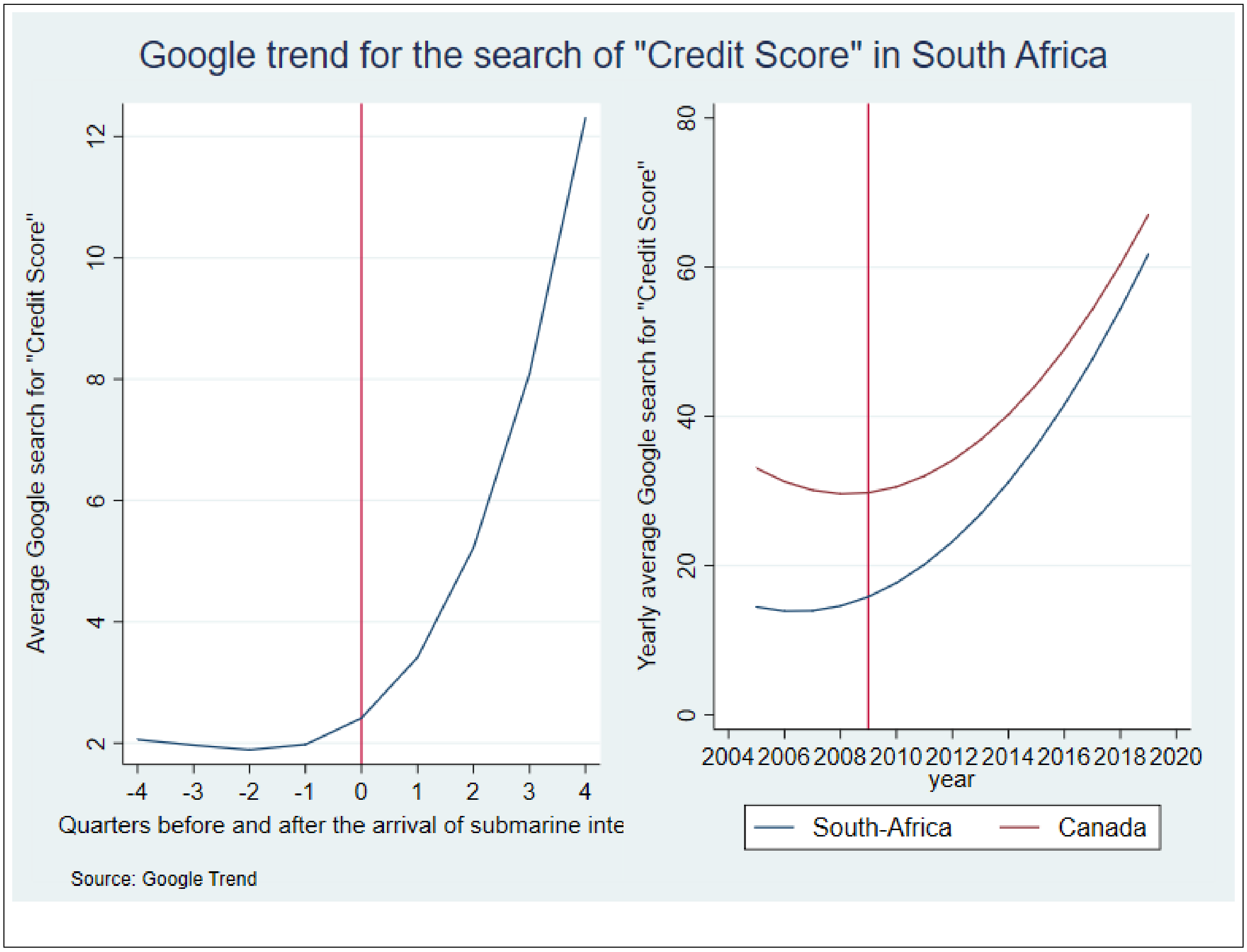


Figure 2-13: The Arrival of Submarine Internet and the Trend of Google Search for "Credit Score".

Table 2-2: LSMS Years and Country

Countries			**Surveys Years**				**First Connected**
Malawi		2005	2010	2013	2017	2019	2010
Niger					2011	2014	2012
Tanzania	2008	2010	2011	2012	2014	2016	2009
Uganda		2005	2009	2010	2011	2013	2009

Table 2-3: WBES Years and Country

Countries		**Surveys Years**		**First Connected**
Ghana		2007	2013	2010
Kenya	2007	2013	2018	2009
Mauritania	2006	2014	2015	2012
Nigeria	2007	2009	2014	2010
Senegal	2003	2007	2014	2010
Tanzania	2006	2013	2015	2009

Table 2-4: Balance Check: Characteristics prior to Submarine Internet Arrival

Variables	Mean for connected areas (Std.Dev)	Mean for non-connected areas (Std.Dev)	Mean difference (t-stat)
	LSMS		
Applied for bank credit	0.1133	0.0656	-0.0477
	0.317	0.2476	(-23.7111)
Bank in community	0.8242	0.0881	-0.7361
	0.3806	0.2835	(-340.554)
Got a bank credit	0.811	0.2167	-0.5944
	0.3915	0.412	(-207.8222)
Credit source of start-up capital	0.0847	0.0776	-0.0071
	0.2785	0.2676	(-3.3885)
Better access to agricultural credit advices	0.429	0.1938	-0.2352
	0.4949	0.3953	(-82.43)
Better access to non-agricultural credit	0.1207	0.8014	0.6807
	0.3258	0.3989	296.2144
Better access to agricultural credir	0.4042	0.1895	-0.2147
	0.4907	0.3919	(-75.8853)
Services	0.5521	0.7902	0.2381
	0.4973	0.4072	(-86.9543)
Trading	0.0943	0.138	0.0436
	0.2923	0.3449	(-20.33227)
Farming	0.1626	0.5728	0.4102
	0.369	0.4947	(-115.7349)
Services business owner	0.0534	0.1003	0.0469
	0.2248	0.3004	-21.5863
Has a permanent position	0.4042	0.3439	-0.0603
	0.4907	0.475	(-17.8896)
Hours worked in the last seven days	19.4713	18.1129	-1.3584
	10.6285	7.1435	(-21.34989)
Run its own business	0.6588	0.2965	-0.3622
	0.4741	0.4567	(-126.8506)
Loan was repaid	0.4432	0.4663	0.0231
	0.4968	0.4989	-5.4924
	WBES		
Ability to provide collateral	0.6413	0.8000	0.1587
	0.4887	0.4000	(7.0152)
Has a bank line of credit	0.6812	0.4145	-0.2667
	0.4796	0.4974	(-9.9417)
Better financing access	0.4532	0.5225	0.0693
	0.574	0.5000	(2.2825)
Percentage of capital borrow from bank	6.5923	12.0112	5.4189
	17.9474	23.1117	(4.1867)
Establishment has a website	0.2142	0.2878	0.0736
	0.4314	0.4757	(2.9778)
Telecommunication is an obstacle to operations	1	0	-1
	0.4274	0.4717	(-2.7585)
Internet is source of information to find new suppliers	0.0173	0	-0.0173
	0.2267	0.3874	(3.8714)

Table 2-5: Descriptive Statistics

Variable	N	Mean	Std.Dev	Min	Max
LSMS					
Distance to optical fiber nodes	464,858	59.977	110.3533	0.1762	1105.633
Has electricity	464,858	0.1089	0.3115	0	1
Fast Internet	464,858	0.0389	0.1934	0	1
Run its own business	464,858	0.4315	0.4953	0	1
Bank in community	348,293	0.0777	0.2677	0	1
Financial credit, source of start-up capital	213,101	0.1099	0.3128	0	1
Better access to agricultural credit	322,949	0.2471	0.4313	0	1
Better access to non-agricultural credit	322,949	0.1555	0.3624	0	1
Better access to agricultural credit advises	322,949	0.2577	0.4374	0	1
Secondary education or higher	464,858	0.2681	0.443	0	1
Has permanent position	323,760	0.2900	0.4538	0	1
Hours worked in last seven days	207,770	17.7173	5.4937	1	108
Loan was repaid	96,547	0.6157	0.4864	0	1
age	464,858	40.562	10.9033	20	64
Female individual	464,858	0.5317	0.4990	0	1
Rural residence	464,858	0.8025	0.3981	0	1
Survey year	464,858			2005	2019
WBES					
Distance to optical fiber nodes	15,028	21.3146	62.1787	0.149423	702.5198
Fast Internet	15,028	0.4028	0.4905	0	1
Ability to provide collateral	14,017	0.8192	0.3849	0	1
Has a bank line of credit	14,017	0.7038	0.4566	0	1
Better financing access	11,419	0.4511	0.4976	0	1
Percentage of capital borrowed from bank	14,242	6.0484	16.0151	0	100
Establishment has a web site	13,849	0.2003	0.4003	0	1
Telecommunication is an obstacle to operations	8,501	0.7294	0.4443	0	1
Internet is source of information to find new suppliers	2,500	0.1	0.3001	0	1
Complex credit application procedure	7,143	0.1412	0.3879	0	1
National sales as percentage of annual sales	14,372	92.3066	22.3117	0	100
Direct exports as percentage of annual sales	13,397	0.0421	0.1567	0	1
Losses as percentage of annual sales	13,906	22.0324	28.8619	0	76
Firm productivity	14,372	1.7353	0.6269	0	2.7778
Year	15,028			2003	2018

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

Table 2-6: Average Effect of Fast Internet on Credit (Demand and Supply)

Panel A

	(1)	(2)	(3)
	Applied for Bank Credit		
	Sample of business owners	**Business owner in urban area**	**Business owner in rural area**
Fast Internet	0.0201***	0.0516***	0.00518***
	(0.00196)	(0.00378)	(0.00160)
N	360912	45133	315779
r-sq	0.0970	0.0583	0.108
ymean	0.0289	0.0386	0.0276

Panel B

	(1)	(2)	(3)
	Bank in community		
	All Sample	**Urban area sample**	**Rural area sample**
Fast Internet	0.0456***	0.0408***	0.0868***
	(0.00278)	(0.00427)	(0.00414)
N	348293	42437	305856
r-sq	0.325	0.432	0.341
ymean	0.0762	0.126	0.0692
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. $*** p < 0.01, ** p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-7: Average Effect of Fast Internet on Financial Credit Supply

Panel A

	(1)	(2)	(3)
	Got a bank credit		
Variables	**All Sample**	**Urban area sample**	**Rural area sample**
Fast Internet	0.0112**	0.0489***	-0.0320***
	(0.00446)	(0.00787)	(0.00646)
N	403362	65434	337928
r-sq	0.213	0.263	0.218
ymean	0.272	0.284	0.270

Panel B

	(1)	(2)	(3)
	Credit is source of start-up capital		
	Sample of business owners	**Urban area sample**	**Rural area sample**
Fast Internet	0.0358***	0.0606***	0.0422***
	(0.00749)	(0.00963)	(0.0135)
N	213101	45776	167325
r-sq	0.147	0.148	0.168
ymean	0.115	0.103	0.119
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. $^{***}\ p < 0.01, ** \ p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-8: Average Effect of Fast Internet on Financial Credit Supply (Business Owners)

	(1)	(2)	(3)	(4)
	Got a bank credit			
Variables	All sample	Sample of business owners	Sample of business owners of urban area	Sample of business owners of rural area
Fast Internet	0.0112**	0.0513***	0.0695***	-0.0886***
	(0.00446)	(0.00687)	(0.00929)	(0.0103)
N	403362	230161	47389	182772
R-sq	0.213	0.252	0.317	0.247
ymean	0.272	0.282	0.325	0.299
Age	✓	✓	✓	✓
Gender	✓	✓	✓	✓
Place of residence	✓	✗	✗	✗
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-9: Average Effect of Fast Internet on Financial Credit Supply (Enterprises Surveys)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	World Bank Enterprises Surveys (WBES)							
Variables	Abilitty to provide collateral	Abilitty to provide collateral	Has a bank line of credit	Has a bank line of credit	Better financing access	Better financing access	Percentage of capital borrowed from bank	Percentage of capital borrowed from bank
Fast Internet	0.0313***	0.0434***	0.0579***	0.0692***	0.0519***	0.0582***	2.7160***	2.2238***
	(0.00949)	(0.0098)	(0.0105)	(0.0109)	(0.0157)	(0.0162)	(0.3516)	(0.3488)
N	14017	13906	14017	13906	11419	11358	14242	13746
R-sq	0.067	0.069	0.317	0.322	0.070	0.071	0.0042	0.026
ymean	0.865	0.865	0.695	0.695	0.451	0.450	6.048	6.0487
Enterprise's city connected	✗	✓	✗	✓	✗	✓	✗	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if a firm is connected to Internet (i.e. is located in a city in which at list one backbone network passes through) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-10: Average Effect of Fast Internet on Agricultural/Non-Agricultural Credit

	(1)	(2)	(3)	(4)
	Got a bank credit			
Variables	**Owners of non-agricultural business in urban area**	**Owners of non-agricultural business in rural area**	**Owners of agricultural business in urban area**	**Owners of agricultural business in rural area**
Fast Internet	0.0347**	0.112***	0.0619***	-0.169***
	(0.0140)	(0.0265)	(0.00894)	(0.0119)
N	15962	43427	31427	139345
R-sq	0.416	0.452	0.356	0.255
ymean	0.196	0.200	0.315	0.336
Age	✓	✓	✓	✓
Gender	✓	✓	✓	✓
Place of residence	✓	✗	✗	✗
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-11: Average Effect of Fast Internet on Agricultural/Non-Agricultural Credit

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Got a bank credit								
Variables	Owners of non-agricultural business	Owners of non-agricultural business in urban area	Owners of agricultural business in rural area	Owners of agricultural business	Owners of agricultural business in urban area	Owners of agricultural business in rural area	Owners of services business	Owners of services business in urban area	Owners of services business in rural area
Fast Internet	0.0295**	0.0347**	0.112***	-0.0333***	0.0619***	-0.169***	0.0559***	0.0269***	0.0105
	(0.0119)	(0.0140)	(0.0265)	(0.00635)	(0.00894)	(0.0119)	(0.00717)	(0.00888)	(0.0173)
N	59389	15962	43427	170772	31427	139345	133268	36751	96517
R-sq	0.428	0.416	0.452	0.247	0.356	0.255	0.320	0.346	0.341
ymean	0.199	0.196	0.200	0.332	0.315	0.336	0.246	0.211	0.259
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✗	✗	✓	✗	✗
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-12: Average Effect of Fast Internet on Economic Activities in Rural Areas

	(1)	(2)	(3)	(4)
	Farming is main occupation		**Trading in main occupation**	
Variables	**Sample of workers of rural area**	**Owners of business in rural area**	**Sample of workers of rural area**	**Owners of business in rural area**
Fast Internet	-0.0325***	-0.0101*	0.0509***	0.0714***
	(0.00448)	(0.00530)	(0.00817)	(0.0104)
N	316081	162359	383583	187860
R-sq	0.330	0.379	0.272	0.293
ymean	0.943	0.942	0.409	0.418
Age	✓	✓	✓	✓
Gender	✓	✓	✓	✓
Place of residence	✗	✗	✗	✗
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-13: Average Effect of Fast Internet on Economic Activities in Rural Areas

	(1)	(2)	(3)	(4)	(3)	(4)
	Farming is main occupation		**Trading in main occupation**		**Services in main occupation**	
Variables	**Sample of workers of rural area**	**Owners of business in rural area**	**Sample of workers of rural area**	**Owners of business in rural area**	**Sample of workers of rural area**	**Owners of business in rural area**
Fast Internet	-0.0325***	-0.0101*	0.0509***	0.0714***	0.0428***	0.0746***
	(0.00448)	(0.00530)	(0.00817)	(0.0104)	(0.00738)	(0.00454)
N	316081	162359	383583	187860	383583	161537
R-sq	0.330	0.379	0.272	0.293	0.355	0.154
ymean	0.943	0.942	0.409	0.418	0.467	0.123
Age	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓
Place of residence	✗	✗	✗	✗	✗	✗
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the country level. **** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-14: Average Effect of Fast Internet on Financial Credit Accessibility

Panel A

	(1)	(2)	(3)
	Better access to agricultural credit		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.0413***	0.0264***	-0.131***
	(0.00480)	(0.00696)	(0.00788)
N	322949	57569	265380
R-sq	0.201	0.238	0.213
ymean	0.255	0.226	0.261

Panel B

	(1)	(2)	(3)
	Better access to non-agricultural credit		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.0777***	0.0229***	0.150***
	(0.00409)	(0.00593)	(0.00725)
N	322949	57569	265380
R-sq	0.239	0.280	0.252
ymean	0.160	0.196	0.153

Panel C

	(1)	(2)	(3)
	Better access to agricultural credit advices		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.00298	0.0256***	-0.0264***
	(0.00481)	(0.00837)	(0.00696)
N	322949	57569	265380
R-sq	0.209	0.238	0.220
ymean	0.266	0.226	0.274
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-15: Aggregate Effect of Fast Internet on Credit to the Private Sector and on Banks Excess Liquidity

	(1)	(2)	(3)
		Country level	
Variables	**Number of commercial bank branches**	**Domestic credit to private sector as % GDP**	**Bank liquidity reserves to bank assets ratio**
Fast Internet	1.942***	4.897***	-4.243**
	(0.335)	(1.463)	(1.996)
N	429	531	555
R-sq	0.773	0.953	0.613
ymean	3.275	18.330	20.029
GDP	✓	✓	✓
Country FE	✓	✓	✓
Year FE	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. Fast internet is the interaction between a dummy variable indicating if a country is connected to Internet (i.e. at least one internet secured server exists in the country) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-16: Average Effect of Fast Internet on Firms' Access to Technology and Information

	(1)	(2)	(3)	(4)
	World Bank Entreprise Survey (WBES)			
Variables	**Establishment has a website**	**Telecommunication is an obstacle to operations**	**Internet is source of information to find new suppliers**	**Complex credit application procedure**
Fast Internet	0.0383***	-0.0695***	0.0699**	-0.0329**
	(0.0103)	(0.0167)	(0.0321)	(0.0143)
N	13849	8501	2500	7143
R-sq	0.143	0.0397	0.0820	0.0107
ymean	0.202	0.729	0.100	0.137
Enterprise's city connected	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. Fast Internet is the interaction between a dummy variable indicating if a firm is connected to Internet (i.e. is located in a city in which at list one backbone network passes through) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-17: Average Effect of Fast Internet on Firms' Productivity

	(1)	(2)	(3)	(4)
	World Bank Entreprise Survey (WBES)			
Variables	**National sales as percentage of anuual sales**	**Direct export as percentage of anuual sales**	**Losses as percentage of anuual sales**	**Firm productivity**
Fast Internet	3.390***	0.00448	-5.266***	0.192***
	(0.518)	(0.00410)	(0.853)	(0.0153)
N	14372	13397	13906	14372
R-sq	0.0181	0.0655	0.0223	0.0317
ymean	92.33	0.0379	23.31	1.735
Enterprise's city connected	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if a firm is connected to Internet (i.e. is located in a city in which at list one backbone network passes through) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-18: Average Effect of Fast Internet on Employment and Entrepreneurship

Panel A

	(1)	(2)	(3)
	Has a permanent position		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.0415***	0.0271***	0.131***
	(0.00481)	(0.00696)	(0.00788)
N	323760	57613	266147
R-sq	0.202	0.245	0.212
ymean	0.265	0.238	0.270

Panel B

	(1)	(2)	(3)
	Hours worked in the last 7 days		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	1.882***	3.999***	0.291***
	(0.164)	(0.254)	(0.0743)
N	207770	27137	180633
R-sq	0.102	0.141	0.0945
ymean	17.68	19.85	17.35

Panel C

	(1)	(2)	(3)
	Run its own business		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.0265***	0.0179***	0.0409***
	(0.00442)	(0.00588)	(0.00796)
N	464858	80436	384422
R-sq	0.220	0.201	0.240
ymean	0.511	0.612	0.490
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-19: Average Effect of Employment Status on Credit Access

Panel A

	(1)	(2)	(3)
	Got a bank credit		
Variables	**All sample**	**Urban area**	**Rural area**
Has a permanent position	0.884***	0.859***	0.897***
	(0.000853)	(0.00240)	(0.000894)
N	322949	57569	265380
R-sq	0.813	0.795	0.829
ymean	0.310	0.302	0.312

Panel B

	(1)	(2)	(3)
	Credit is source of start-up capital		
Variables	**Sample of business owners**	**Business owners in urban area**	**Business owners in rural area**
Has a permanent position	0.441***	0.508***	0.430***
	(0.00198)	(0.00476)	(0.00220)
N	237508	49261	188247
R-sq	0.445	0.510	0.447
ymean	0.114	0.103	0.117
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

*Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Table 2-20: Average Effect of Fast Internet on Loan Repayment

	(1)	(2)	(3)
	Loan was repaid		
Variables	**All sample**	**Urban area**	**Rural area**
Fast Internet	0.0173**	0.0441***	0.0098**
	(0.00824)	(0.0121)	(0.0072)
N	96547	14183	82364
R-sq	0.212	0.267	0.233
ymean	0.616	0.576	0.623
Age	✓	✓	✓
Gender	✓	✓	✓
Place of residence	✓	✗	✗
Year FE	✓	✓	✓
Country FE	✓	✓	✓
Grid connect to electricity FE	✓	✓	✓

Note: Robust standard errors in parentheses, clustered at the country level. *** $p < 0.01, **p < 0.05, *p < 0.1$. *Fast Internet is the interaction between a dummy variable indicating if an individual is connected to internet (i.e. lives within 2 kilometers of a fiber node) and an indicator variable for whether a backbone network in their country is connected to at least one submarine cable.*

Chapter 3

Internet, Roads, and Employment: Evidence of Infrastructure Complementarity in Sub-Saharan Africa

3.1 Introduction

Several works in the economic literature illustrate the capital role that investments in infrastructures play on economic growth through the job creation process. Duflo et al. (2015) exhibited the causality between the nearness to roads networks and the increase of the GDP per capita within regions in China. Atack et al. (2009) found that the construction of railroads had induced economic growth in American Midwest by triggering the urbanization. Hjort & Poulsen (2019b) revealed that submarine internet impacts employment rates in a positive way and in proportion to the skill level in Sub-Saharan African countries. However, most of these works point to the isolated impact of different types of infrastructure on employment. Extraordinarily few works have investigated the joint effects of infrastructures.

Moneke (2019b) showed that the road network and electrification complement each other to improve Ethiopian welfare by 11 percent, while the same infrastructure's isolated effects on well-being are only 2 percent. In this paper, we study a different complement to road infrastructure. More precisely, we are investigating how the arrival of the submarine internet complements the effect of the road network to influence employment possibilities in Sub Saharan African countries. As far as we know, this work is the first to explore the joint influences of roads and Internet on employment, the structure of the labour market, and on private and public entrepreneurship initiative.

Internet connectivity has rapidly expanded since the 1990s. Roughly half the world today uses the internet, including over one-third of Africans and nearly a quarter of South Asians - while access has accelerated in low-income countries (LICs). Worldwide there were 17 broadband subscriptions per 100 people in 2020.[1]

Many countries in the developing world are making significant investments in Internet infrastructure, hoping that connectivity can facilitate and create more and better jobs. While technology optimists see digitization as an opportunity for low-income developing countries to leapfrog into the 21st century, others emphasize the disruptive effects digital technologies are expected to have on labor markets.[2] In some contexts, Information and Communication Technology (ICT) could reduce longer-term job growth through automation or a shift in activity away from manufacturing (Rodrik, 2016). Africa needs more than anywhere else in the World to create more and better jobs. Hjort & Poulsen (2019a) were among the first to empirically take up the question of whether Internet connectivity affect jobs in Africa. The authors

[1]Source: International Telecommunication Union (ITU) World Telecommunication / World Bank ICT Indicators Database https://data.worldbank.org/indicator/IT.NET.BBND.P2

[2]https://blogs.worldbank.org/developmenttalk/moonshot-africa-and-jobs

provide insight into the mechanisms through which fast Internet increased employment. Our paper investigates whether complementary investments in transport infrastructure to lower trade costs may work in tandem with modern ICT technologies to boost job creation and incomes.

To our knowledge there is no paper on the complementarity or substitutability between access to roads and access to Internet connectivity, to complement works on roads and electricity by Moneke (2019b) and Abbasi et al. (2022). Both roads and Internet access provide better market access to firms and consumers. Access to Internet connectivity can substitute for access to roads, for example; working from home reduces the use of roads to commute. However, access to Internet can only support e-commerce in rural and remote areas if good roads allow for trade from these regions. Similarly, Internet may reduce information frictions for firms to use better inputs, access better workers, or train their workers on the job; but good roads remain necessary to import these inputs, allow workers to easily come to work, or have the incentives to improve production through labor productivity investments. This chapter aims to identify whether roads and Internet access are substitutes or complements for job outcomes in Africa, and whether these impacts vary across genders, rural and urban locations, skills and occupations.

We exploit the geographic positions of households and firms (from the Demographic and Health Surveys (DHS), the World Bank Enterprises Surveys (WBES) and the Living Standard Measurement Study (LSMS)), the localisation of roads[3] and optical fiber nodes[4] of sub-Saharan African countries, to examine how the development of roads networks complements/substitutes the access to fast Internet to encourage entrepreneurship, and therefore strengthen job creation. Using a combination of the difference-in-differences technique and instrumental variables, we estimate the joint effects of a change in the proximity to major roads and accessibility to ICT infrastructures on the labour market.

We find that an individual who gets 10 kilometers closer to the nearby major road, and who has access to Internet sees his/her odds of working increase by 68.4 percentage points, while another one, not connected to Internet and for who the distance to the closest main road has decreased by 10 kilometers will have his odds to be working increase by only 7.84 percentage points. Similarly, a person connected to Internet who has not experienced any improvement in his accessibility to the road network will be only 5.68 percent more likely to be working. On the firms side, we find that the simultaneous provision of both types of infrastructure increases the demand for labor force by businesses; in particular by allowing for an increase in the number of hours of work per week. More specifically, we find that

[3]Data on roads are those constructed by Jedwab & Storeygard (2017)

[4]Data on optical fiber nodes are from Africa Bandwidth Maps and the submarine Internet's maps of Mahlknecht (2014)

companies that are connected to high-speed Internet, when they get closer to the road by 10 kilometers, their number of weekly operating hours are higher by 1 hour and 33 minutes compared to that of companies that have experienced an improvement in access to only one of the two infrastructures. This result is consistent with another of our findings that, firms with improved access to both infrastructures have more full-time employees than firms with less or limited access to either infrastructures.

Our findings imply that both infrastructures complement each other to improve the labour demand. Indeed, we find that the likelihood of an individual to have been employed during the last twelve months is 8.33 percent higher when the individual has access to the Internet but is located at an unchanged distance to the nearby major road. However, that same individual will be 10 percent more likely to have been employed in the year preceding the interview, when he has access to Internet and has gotten one kilometer closer to the nearest major road. Expressly, the impact of Internet on the labour demand is enhanced by 1.67 percentage points when the access to Internet is supplemented by a better proximity to the road's network.

We identify two main channels through which the joint provision of means of transport and ICT affect the job market: (1) the private sector entrepreneurial initiatives and (2) the public sector entrepreneurial initiatives. We find that the provision of both infrastructures simultaneously increases the demand for labor by the public sector by 6.1 percentage points in comparison to the case where only one of the two infrastructures is provided. Moreover, people for whom access to the two infrastructures has been improved marginally have 15.5 percentage points greater chance of running a business of their own, compared to those whose access to only one of the infrastructures has been improved.

Contributions to the literature This article adds to the vast literature on the impact of infrastructure's expansion on economic growth and development. Some studies have shown that the construction of highways networks improves welfare by a percentage considerably greater than its cost (Jedwab & Storeygard (2017); Ejaz Ghani (2016); Allen N. et al. (2001); Shatz et al. (2011); Donaldson (2010); Aschauer (1989)). Some other studies showed the impact of better accessibility to roads or railroads on per capita GDP (Duflo et al. (2015); Donaldson (2010); Atack et al. (2009)). Some others assessed the effects on the short-run employment opportunities (Estache et al. (2013)). All of these papers have focused solely on the isolated effect of the road on economic variables. Moneke (2019b)'s work is one of the few studies which investigates the joint impact of road and electrification on employment in Ethiopia. Abbasi et al. (2022) add to this preliminary literature by studying the heterogeneities in the joint effect of road and electrification in Sub-Saharan Africa. Our study is different in that we are interested in the interaction between investment in road construction and the arrival of the submarine internet.

Existing works have shown that investing in information and communication technologies impacts economic expansion in several ways: it promotes labour demands (Hjort & Poulsen (2019b) and increases productivity (Hjort & Poulsen (2019a);Goldfarb & Tucker (2019); Draca et al. (2006), Rollo & Paunov (2015)); it reduces transaction costs, since Internet provides access to information, and so reinforces structural innovations (T. Allen (2014); Brossard et al. (2007)); it facilitates innovations (Bertschek et al. (2019);Rollo & Paunov (2015)); enhances competition, and leads to new techniques of production, improvement in trades and consumption (Grazzi & Jung (2019);Miller & Atkinson (2014);Waverman et al. (2005)). On one hand, our work aims to link the aforementioned stimulus to the emergency of new employment opportunities through entrepreneurship. On the other hand, note that all those studies focus on the isolated impacts of the internet and other infrastructures. Our work therefore stands out from these by analyzing the additional impact arising from the complementarity between infrastructures on employment opportunities.

Our other contribution to the literature is the development of two original instrumental variables to address the possible endogeneity of infrastructures' locations. We exploit the least-cost path procedure to construct theoretical roads' networks connecting the centroids of different ethnic territories in sub-Saharan Africa.[5] We interact the resulting exogenous distance to the nearest main road with the accessibility to submarine internet to show that they complement each other to boost the private and public entrepreneurship, which in turn generates employment opportunities, shifts the labour force from low skilled to higher skilled jobs, and strengthens job stability.

The remainder of this paper is organized as follows. The next section presents the conceptual framework. Section 3 describes the context, the data and our empirical strategy. Sections 4 and 5 presents the main findings. It also presents findings on the heterogeneity of infrastructure complementarity by key socioeconomic and demographic characteristics. Section 6 concludes.

3.2 The Gains from Road and Internet Connectivity

Following Hjort & Tian (2021), we present the potential mechanisms driving the job and economic impacts of the simultaneous provision of Internet connectivity and roads' network. Internet and road access can drive economic development through their impacts on both the supply-side and the demand-side of an economy.

[5] We exploit data on land slope to build the series of accumulated costs of moving from one ethnic location to another one; the optimal path is the one that minimizes the construction of the link connecting two ethnic homelands.

3.2.1 Supply-side

Internet and roads can directly affect the productivity of firms, workers, and other inputs in the production process. Findings from Z. Li et al. (2017) suggest significant and positive returns of road investments in China in terms of manufacturing firms' productivity. A similar study by Gordon & Li (1995) regarding government's firms reveals a 4.6 percent increase in their productivity, generated by roads' investments, amounting to 3.6 annual total factors' productivity (Zhu (2012)). Gibbons et al. (2016) obtained similar findings for the case of the UK; the authors found new roads infrastructures greatly impact Britain firms' productivity. As for the accessibility to Internet connectivity, it can directly affect workers' job productivity, induce human capital accumulation, and changes firm-worker matching.

First, several empirical papers point towards a direct impact of road accessibility on wages and employment (Gibbons et al. (2016); Yamauchi (2016); Matas et al. (2015); Yamauchi et al. (2011); Frish & Tsur (2010); Reardon et al. (2001). Other empirical studies also point out the direct influence of Internet connectivity on employment and wages. Chen et al. (2020) examine a policy reform in China around 2000 that increased Internet speeds and document significant increases in workers' wages and firms productivity in response to the Internet-upgrading program. In the context of Brazil, Almeida et al. (2017), POLIQUIN (2020), and Tian (2021) exploit a roll-out of new internet infrastructures and find that broadband access increases workers' wages on average. As an exception, Dutz et al. (2017) find a negative correlation between increased Internet access in Brazil and average wages. In Nigeria, Bahia et al. (2020) examine how the roll-out of mobile broadband affects labor market outcomes, household consumption, and poverty in Nigeria. Using a difference-in-differences approach tracking individual households, they show that in the Nigerian context, Internet connectivity increases labor force participation and employment.

Second, another strand of the literature also points towards a direct impact of Internet connectivity on human capital development. First, Internet connectivity can facilitate on-the-job training. Evidence that connected firms invest more in their workers' human capital has been found in Hjort & Poulsen (2019a) for six African countries and in Mouelhi (2009) for Tunisia. Second Internet connectivity can affect human capital development at home and in school. Bianchi et al. (2020) show that connecting high-quality teachers in urban areas of China with million students in rural primary and middle schools improved students' long-run academic achievement, labor market outcomes, and Internet use. Overall, the evidence on the impacts of connected schools on test scores is mixed. Evidence of positive impacts has been provided in the context of Peru Kho et al. (2018) and Malawi Derksen et al. (2019). In contrast, Malamud et al. (2019) for Peru and Bessone & Dahis (2020) for Brazil find no impact on test scores.

Finally, Internet can affect labor productivity by improving firm-worker matching, especially in developing economies where frictions on the labor market are large. Empirical evidence can be found mostly for high-income countries; see Kuhn & Mansour (2011) for the US, and Bhuller et al. (2020) for Norway. Lederman & Zouaidi (2020) find a robust negative relationship between Internet usage and long-term frictional unemployment across countries. Internet can also increase firm level productivity by facilitating adoption of other technologies, like tangible inputs such as machines and intermediate materials, or intangible inputs such as management, organizational practices, and services. Evidence on take-up of tangible inputs in response to Internet connectivity is limited, but Internet connectivity has been shown to affect the organization of production and trade in developing countries. Houngbonon et al. (2022) show evidence that individual firms in Africa are 20 and 12 percentage points more likely to undertake respectively process and product innovation when fast internet becomes available. Tian (2021) shows that internet access allows firms in urban areas to reorganize production to enhance collaboration and facilitate division of labor.

Using enterprise level panel data for Britain, Gibbons et al. (2016) examined the effects of investment in roads' construction on employment and labour productivity in the UK. They found that improvement in the accessibility to road substantially increases workers productivity at the individual level. In China, Z. Li et al. (2017) found that road infrastructure investment yields an annual return of 11.6 percent from the productivity gain. In parallel, using firms' portfolio data-files, H. Li & Li (2013) found that investing in roads, for the case of China, has caused an increase in firms' productivity, with an annual rate of return close to 10 percent.

3.2.2 Demand-side

While roads can impact employment and firms' expansion through exports and trade (Volpe Martincus et al. (2017); Mbekeani (2010)), through regional integration (Mbekeani (2010)), or through new market entry enhancement (Melo et al. (2010); Holl (2004)), Internet connectivity can also affect economic activity by directly expanding firms', workers' and consumers' market access, and by addressing information frictions. For example, e-commerce may allow firms to make their products accessible to more consumers, especially in rural and remote regions. Internet connectivity appears to enable firms to expand their sales between regions within countries (see Fan et al. (2018) for China) and through exporting and importing (see Hjort & Poulsen (2019a) for Africa). Internet appears to lower the prices and expand the variety consumers face (see Couture et al. (2021) for China). In contrast to barriers to accessing markets, information frictions are pervasive in developing countries. Internet access can lead to (1) a reduction in price dispersion, (2) higher local prices when buyers have a degree of monopsony power (see Goyal (2010) for India, and Guerrero Barreto & Ritter Burga (2014) for Peru), (3) a reduction in uncertainty about product quality, and (4) improvement in communication with trade partners (see Leuven et al. (2018) and Fernandes et al. (2019) for China).

3.2.3 Heterogeneities

Important heterogeneity in Internet's wage and productivity effects across demographic groups have been uncovered in the literature. Many papers have found that Internet connectivity appears to especially benefit female workers. In Vietnam, Chun & Tang (2018) find suggestive evidence that firms that increase their ICT use also increase their female labor share. In Brazil, Dutz et al. (2017) find that employment growth from internet arrivals in different areas of Brazil is greater among low-skilled female-filled jobs. Juhn et al. (2013) document a similar impact for Mexico, and Bahia et al. (2020) for Nigeria.

The evidence on Internet technology's skill bias in developing countries is mixed. Khanna & Sharma (2018) show descriptive evidence of complementarity between ICT and non-routine tasks. Chen et al. (2020) find that the adoption of high-speed Internet benefited Chinese firms in more skill-intensive industries and with more educated workers. Dutz et al. (2017) also find evidence that, within the manufacturing sector in Brazil, Internet access appears to raise wages in medium- and high-skill jobs, but not in low-skill jobs. In Tanzania, Bahia et al. (2021) show evidence that broadband availability induces increased labor force participation and wage employment among young, educated men. Hjort & Poulsen (2019a)) show that the gradual arrival of fast Internet infrastructure in Africa increases employment rates even for less- educated worker groups, although the estimates are considerably larger for more educated workers.

The evidence on differences across urban and rural areas in developing countries is also mixed. Masaki et al. (2020) find that the labor market impact of fast internet in Senegal is larger among households in urban areas. A few other studies find especially beneficial impacts in rural areas (see Bahia et al. (2020) for Nigeria).

The effects of road construction have been shown to vary depending on location, industry type, and level of qualification. Baum-Snow et al. (2017) provided evidence that new regional highways promote concentration of outputs but also concentration of the working-age population into regional prefectures in China. Moneke (2019b) found that investment in road expansion improves productivity as isolated areas get closer to the road network, with some manufacturing sectors growing to be profitable for exports. The study also found that road access alone causes a large shift of employment from the manufacturing sector to high-skill activities, and, by contrast, causes service employment to rise, largely in informal, small-size retail businesses.

3.3 Context, Data, and Empirical Strategy

In this section, we describe the context of infrastructures in sub-Saharan Africa, outline the data sets we use to show the causal join effect of investment in roads and the Internet, and present the empirical strategy.

3.3.1 Context

The literature documenting the impact of infrastructure is still preliminary in the context of Sub-Saharan Africa. In this region, the infrastructure deficit is still very significant. The African Development Bank together with the Africa Infrastructure Consortium asserts that less than 20 percent of the population have access to electricity; barely a third of residents of rural areas live near a road. This picture of African's infrastructures is indeed very comparable to that of our database (to be described below), where less than 25 percent of our sample is connected to the road network, with an average of 15.19 kilometers separating households and the nearest road in our LSMS sample and 9.70 kilometers in our DHS sample.

The state of infrastructures in sub-Saharan Africa reveals at least two important facts. First, there is a great shortfall in relation to its needs. Indeed, the sub-Saharan Africa population is growing faster than the level of economic growth and development (and implicitly job creation). As we can see in figure (3-1), the populations of Cameroon and Malawi are increasing at a constant speed, while the employment rate shows a downward trend, meaning that population is growing faster than employment opportunities. As a result, the unemployment rate could be higher in the future Birdsall (1977). However, it is interesting to note that Singapore[6], although presenting a population growth rate very close to that of Cameroon or Malawi, performs very well to maintain an upward trend in its employment rate. This explains the reason for identifying the means that can enable the countries of sub-Saharan Africa to improve their job creation process, especially the role of infrastructures.

Second, the development of infrastructures in sub-Saharan Africa lags considerably behind that of other poor regions. Therefore, the high transport costs due to infrastructure problems, or the cost of access to information (not necessarily accessible to all) limits the efficiency of African markets, and thus constitute a brake on the labour demand. Indeed, if we take the case of Cameroon and Singapore which, in 1994, had the same level of internet access, as

[6]Singapore and Cameroon had similar levels of development in 1994.

shown in figure (3-2), it is interesting to note the significant gap that exists between these two countries at the entrepreneurship level, although these two countries both exhibit an annual population growth rate stable at around 2 percent.

The conclusion is exactly the same when we compare Cameroon and Malawi, which are both countries in sub-Saharan Africa. Although they both have Internet accessibility showing an upward trend, the gap between these two countries explains the gap they maintain in terms of their level of entrepreneurship and therefore job creation.

3.3.2 Data Description

Our data come from the following sources:

Demographic and Health Surveys (DHS) The DHS are cross-sectional surveys that have been conducted in the majority of developing and middle-income countries since the 1980s. They are representative at the national and subnational levels. All household members' demographic and socioeconomic status information (e.g., age, gender, education, occupation...) are collected during the interviews. Furthermore, selected women and men provide information on a variety of other variables, such as health status, fertility history, migration, and children. The DHS Program has gathered geo-reference data (longitude and latitude) and other relevant information such as altitude in some surveys, allowing these surveys to be combined with a wide range of external data and information. Recent studies have used DHS to examine the economic and social impacts of infrastructure in developing countries (Okoye et al., 2019; Hjort & Poulsen, 2019a; Moneke, 2019a; Canning et al., 2020; Herrera Dappe & Lebrand, 2021; Lebrand, 2022; Abbasi et al., 2022).

We rely on data from 59 surveys conducted in Sub-Saharan African countries. The total sample size is 806,378; with 552,022 women and 254,356 men. Grid cells of sized 20 km by 20 km are constructed around each location to account for any time invariant factor that might affect our infrastructure location and our outcome variables. The characteristics of individuals in the sample are summarised in table (3-2).

We divide employed people into two categories: high-skilled workers and low-skilled workers. In some parts of our analysis, we also differentiate between agricultural workers and non-agricultural workers. Employed persons are considered skilled workers if they operated in any of the following occupation categories at the time of the survey: professional, managerial, skilled manual labour, services, or agriculture (employee). Individuals are classified as low-skilled workers if they operated in one of the following occupations: domestic, unskilled manual labour, or agriculture (self-employed). In our data, 67.70 percent of respondents claimed to be employed, while 32.30 reported unemployment status.

Skilled occupations are distributed as follows in our database: sales (38 %); manual labour (18 %); agricultural employee (16 %); service-sector (22.9 %); professional, technical, or managerial occupations (10.90 %); clerical occupations (2 %). Eighty percent of unskilled workers reported being self-employed in agriculture, with the remaining 20 % working in domestic work or as unskilled manual labourers. Sixty-two percent of respondents lived in rural areas, while 38 percent lived in urban areas. The ages of those who responded to the survey were as follows: 25 and under (39%), 25 to 50 (57%), and over 50 (57%) (0.2 %)).

We divide the survey data into two categories of development level to examine the differences in the effects of infrastructure investment across countries at various stages of development. Using PPP-adjusted GDP per capita, we categorize countries as less developed or more developed. A country is considered less developed if its PPP-adjusted GDP per capita falls below USD 3,400 in 2017. Around 61 percent of countries in our sample are classified as less developed.

World Bank Enterprise Surveys (WBES). The WBES are countrywide representative samples of businesses or firms from all sectors of activity with at least five workers. The WBES are administered to business owners and highest supervisors of the private sector industries. The surveys include information on enterprises' establishments, sources of funding, assets, structure of workforce, operations and type of activities. We exploit the WBES data from 2003 to 2018, of coastal nations that had survey cycles before and after the landing date of submarine internet. These countries are Ghana, Kenya, Mauritania, Nigeria, Senegal, and Tanzania. We limit our sample to Small and Medium Sizes Enterprises (SMEs) that employ a maximum of 80 individuals.

The WBES provides information on the city of establishment of firms. We use that information to combine the data set with the geo-localisation of cities in Africa. This combined data allows us to compute the indicator of whether the establishment's city is connected to Internet; as well as the distance to the closest major road.

After pooling all observations together, we obtain a sample of 15,033 firms. For an Internet connectivity defined in a radius of two kilometers from the closest optical fiber node, about 78 percent of firms are located in a cities connected to Internet, with 36.45 percent connected to submarine Internet. The average distance between the centroide of each city and the nearby main road is 2.31 kilometers; with treated firms located at an average of 300 meters from the nearby major road. Firms' hours of operation range between 10 to 84 hours, with an average of 35 hours a week. About 26.85 percent of them have reported that telecommunication shortness is an important obstacle to their operations. They employ, on average, 16 full time workers, for an average of 6 female workers. More details on this sample are displayed in table (3-3).

Living Standards Measurement Study (LSMS) The LSMS is the World Bank's household survey program. It is dedicated to improve household survey techniques and enhance the quality of microdata and to well enlighten growth programs. The LSMS project works in partnership with the national statistics offices of its eight affiliated countries[7] in Sub-Saharan Africa to model and execute techniques of multi-subject matter, nationwide representative panel household surveys with a sharp attention on agriculture. The main purpose of the project is to promote modernization and productivity in statistical research on the links between agriculture, socioeconomic conditions, non-agricultural activities, and poverty decline. The occurrence of data collection is decided on a country-by-country basis, conditional on data request and the availability of funding. The sampling procedure applied is a stratified two-stage (two-stage probability sample). In the first stage, enumeration areas (EA) are selected with probability relative to the total number of households considered in the EA; in each EA, households are selected, with the purpose to follow and reassess them on their behaviors, welfare, distribution of income, consumption, employment, health, education, and financial decisions. In the second stage, a subset of the main households' sample is taken in identical likelihood in each of the EA. During the following surveys years, households in the same EA are reassessed. However, due to mobility, some households had left the survey and others had entered. The draw of households in the EA is done after a listing operation whose purpose is to update the selection of households.
The LSMS component of our final data set is summarized in table (3-4). In this paper, only 7 out of the 8 countries are considered due to lack of geo-referenced data in the Ivory Coast's sample. Our sample contains 1,577,278 respondents, with ages ranging from 20 to 65 years old. The average population age is 35.8 years old. About 86.2 percent of the sample lives in rural areas. About 52 % of the population is female. They are located on average at 15.2 kilometers from the closest major road, with treated individuals being at 5.6 kilometers from the principal road. Around 5.6 % of individuals in this sample have access to high-speed internet. We present detailed information about this sample in table (3-4).

The **roads dataset** Using geo-coordinates, we link the roads' data to the DHS, WBES and LSMS data sets. Data on roads come from the work of Jedwab & Storeygard (2017). The authors pooled materials from two resources. Originally, Nelson & Deichmann (2004) had produced the localisation of roads existing in 2004 for all Africa; built principally by the US government's Digital Chart of the World database, that contains incomplete information on roads' types. Using the geo-localisation of those roads as a starting point, Jedwab & Storeygard (2017) digitized 64 Michelin road maps constructed between 1961 and 2014 to characterize current road outlook. Roads' types vary among maps, the discrepancy between

[7] Ethiopia, Ivory Coast, Malawi, Mali, Niger, Nigeria, Tanzania, Uganda

highways, other concrete roads, improved roads (laterite or gravel), and dirt roads is universal. Michelin had used the correspondence from road operators, the administrations road maps, explicit information from his tire's stores within Africa; that were combined with the earlier Michelin maps to create the 64 aforementioned road's maps.

Figure (3-3) illustrates the dynamics of the road network of three countries in sub-Saharan Africa over the period 1965 to 2014, where we can note a remarkable evolution between 1976 and 1986.

Indeed, Michelin's merchants collect statistics on projects and the maps are supposed to specify the year when a specific road had been made operational. Michelin's map is in short supply for certain states and periods. We are using grid fixed effects to palliate to this limit. Location along with residents' household are then connected to the transport network. We link the DHS data and this road system using the geocoordinates. In fact, The DHS dataset provides the latitude and longitude of each location. This allows us to estimate the distance between each household and the nearest main road (highway, paved and/or improved road).

In our sample, the average distance to the closest main road is 9.70 km (DHS), 2.31 km (WBES) and 15.2 km (LSMS). In rural areas, this average is 13.43 km against 3.60 km in urban areas in the DHS sample. Looking at the heterogeneity in the proximity to a main road by level of development, in less developed countries, this average is 11.46 km versus 7.50 kilometers in more developed countries.

Internet data We use the Hamilton data from Africa Bandwidth Maps on long-distance optical fibre nodes from 2009 to 2019. The map depicts Africa's terrestrial, satellite, and submarine cable transmission networks in great detail. The Hamilton map includes fibre optic networks (operational, under construction, planned, and proposed), microwave networks (operational and planned), and submarine cables (operational, under construction, planned, or proposed) for over 300 network operators in more than 50 African countries, as well as 72 submarine cable systems. Geographic coordinates of fibre optic nodes are included in the data set. This allows us to calculate the distance between each optical fibre node and the location of individuals. We have information on the year in which each fibre node became operational. Figure (3-4) depicts the evolution of the optical fiber network over time, where we can see a significant increase in the number of optical fiber backbones from the year 2012. In fact, in 2012, a large majority of African countries already have their terrestrial backbones connected to submarines cables.

We also use information from Mahlknecht (2014) on submarine Internet landing points and arrival dates. We combine this data with the location of optical fibre nodes to determine a location connectivity to Internet and to submarine (or high-speed) internet.

When the distance between the centroid of a cluster (enumeration area (for the LSMS) and city (for the WBES)) and the location of the nearest optical fibre node is less than two kilometers, we consider individuals in that location to be connected to the Internet network. In our main sample, approximately 42 percent of people have access to internet. Therefore, an individual is connected to submarine internet if he is connected to an optical fibre node after the arrival of submarine Internet.

3.3.3 Identification Strategy

The objective of this work being to show the joint causal effect of the roads and telecommunication infrastructures on the labor market, we use the distances measured between these infrastructures and the location of our unit of analysis for this purpose. However, the location of each type of infrastructure could be correlated to economic activity. Therefore, we adopt an approach that combines Difference-in-Differences (DID) and Instrumental Variables (IV) techniques. We rely on: (1) a DID strategy to explore the causal effect of Internet and (2) on two IV for the proximity to road and its interaction with Internet, respectively, to explore their causal effects.

The Model

First stage

In the first stage, we estimate the following equation:

$$\begin{aligned} P.Roads_{i,g,t} = \beta_5 + \beta_6 hpy.road_i + \beta_7 Submarine * internet.connected * hpy.road_i \\ + \beta_8 geo_g + \lambda_t + \gamma_g + \delta_c + \varepsilon_i \end{aligned} \tag{3.1}$$

Where $P.Roads_{i,g,t}$ is the measure of the distance between location g of individual i and the nearby main road (or its interaction with the treatment indicator). $hpy.road_i$ is the shortest distance to the conjectural network built from Murdock 1959's map, and $Submarine * internet.connected * hpy - road_i$ is its interaction with the treatment variable. We also control for geographic characteristics (geo_g) including terrain topography and lightening intensity, and for the same fixed effects as in equation (3.2).

Second stage

The simultaneous causal impacts of the accessibility to major roads and the availability of Internet on employment status and other labour's market outcomes are examined by estimating equation (3.2), that combines the two aforementioned strategies:

$$\begin{aligned} X_{i,g,t} = \beta_1 + \beta_2 Submarine * internet.connected_i + \beta_3 Road.Proximity_i + \\ \beta_4 Road.Proximity * Submarine * internet.connected_i + \lambda_t + \gamma_g + \delta_c + \varepsilon_i \end{aligned} \tag{3.2}$$

With X_i, g, t being individual i's employment outcome (or firm i's outcome).
The variable $Road.Proximity_i$ is the proximity to the road. It is a continuous variable representing the negative distance to the nearest main road. The proximity to the nearest main road is measured by taking the length of the shortest straight line between the place of residence and the closest major road. However, given the fact that roads' networks are probably not randomly built across cities, OLS estimates might be biased. Therefore, to correct for that, we use two instrumental variables, the land roughness and the distance obtained from exogenous theoretically built networks.

Submarine is a dichotomous variable taking the value one when a fiber node in a given area is connected to fast internet (submarine's cables). The variable $internet.connected_i$ takes the value one when the grid in which the individual i is located is connected to an optical fiber node. The interaction between these two variables allows to isolate the impact of fast internet on employment status for areas connected to submarine internet from the variations in that same outcome in other localities not connected to fast internet. The variable $Road.Proximity * Submarine * internet.connected_i$, which is the interaction term between the road variable and the binary indicator for fast internet access, captures the joint effect of these two infrastructures on employment status (or occupation).

We also control for years and country fixed effects. We further account for local time-invariant unobserved factors that might bias our estimated effects of infrastructures by controlling for the 20*20 kilometers grid cell fixed effects. We cluster the standard error at the grid cell level.

Who is connected to the Internet?

We define the connectivity to the Internet based on the distance to the closest optical fiber node. The deployment of optical fiber has accelerated considerably in recent years in sub-Saharan Africa, from less that 400 thousand kilometers of network in 2009 to over 1.5 million kilometers in 2019. Optical fiber is a technology with the greatest capacity to transport information. It has the advantage of having a faster speed with less attenuation or loss of signal as one moves away from the network connection nodes. Its high speed and large bandwidth can provide access to high-speed Internet to a large number of individuals. Optical fiber is a particular type of terrestrial backbone cable with the main characteristic of being a very high-speed electrical cable, allowing its signal to be transmitted up to 5 kilometers depending on the type of optical fiber.

Indeed, thanks to its large bandwidth and its higher capacity to get the information circulating very fast, optical fiber cables support extremely high bandwidth and speed; up to 10 Gbps. The amount of information that can be transmitted per unit of optical fiber cable is its most significant advantage. Also, the fact that it offers less signal degradation results in not only a faster Internet connection, but also in an Internet connection that has a large signal transmission range, as shown in figure (3-5). There are exactly two types of optical fiber cables:
(1) The single mode optical fiber, which 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 very high cost, it is mainly used for very long-distance signal transmission.
(2) The most-used type of cable is the multimode optical fiber cable, which has the characteristic of guiding several modes simultaneously, and therefore allows the routing of a greater volume of data simultaneously. It is generally used for short distance signal transmission such as in concessions, campuses or even apartment buildings. Multimode fiber optic cables can transmit a signal of 100 Mbit per second up to 2 kilometers. This justifies why we choose 2 kilometers as the threshold of Internet connectivity in this study. We consider locations positioned within 2 kilometers of the closest optical fiber node as connected locations.

Figures (3-8) shows how employment relates to proximity to central roads' network and access to Internet. To generate these figures, we compute the probability of employment for each given distance to an infrastructure's location, where distance is measured in bins of 10 km. We then fit the prediction of employment on distances. We find that as the distance to a main road increases, the proportion of active population decreases. For example, those living within 10 km of a main road are about 20 percentage points more likely to be participating in the labour market than those living within 100 to 110 km of a main road. A similar pattern is observed regarding the distance to the closest fiber optic. For example, individuals living within 10 km of the closest fiber optic are about 12 percentage points more likely to be working than those living within 90 to 100 km of the closest optical fiber node.

The Causal Effect of Road, Submarine Internet, and their Interaction: IV Approach

Transportation infrastructure was created by colonial rulers to move agricultural products and mineral resources from the interior of countries to export ports Okoye et al. (2019). Infrastructure improvements also served an administrative purpose in each country; better roads and trains allowed colonial authorities to travel to most sections of the country. Despite the fact that the same motivations fueled the creation of transportation networks in the post-independence period, country leaders used a different strategy to expand these networks. Instead, they constructed transportation networks that would enable them to expand the exercise of their authority from the country's "political centre" across their whole territory, including rural locations and all ethnic homelands Herbst (2014); Muller-Crepon et al. (2020).

Therefore, based on the rationale that post-independence African leaders promoted the construction of transportation infrastructures to allow them to govern the entirety of their respective territories from the capital city - the centre of political institutions - and to bolster national integration, we generate a new instrumental variable for proximity to road networks. Construction of transportation infrastructure, particularly roads and trains, has been driven by economic, social, and administrative (political or military) motives in Africa since the colonial era ((Taaffe et al. (1963), Okoye et al. (2019)).

Indeed, the arbitrary partition of Africa by European colonizers at the Berlin Conference in 1884 and 1885 resulted in diverse levels of ethnic diversity among the countries that arose from this experiment. In most of these countries, leaders faced the difficult task of administering a diverse ethnic population while also unifying and integrating them Deng (1997). To promote country-building and national unity, these leaders used a variety of techniques. These measures included promoting the use of a single national language, prohibiting political organizations formed around ethnic groups, and distributing ministerial offices and other public positions among different ethnic groups.

Most African countries used infrastructure development, such as highways, as a tactic for national integration. Infrastructure development as a means of integrating different ethnic groups is not a phenomenon exclusive to African countries. Han & Paik (2017) argued that, "The state has pursued provision of necessary infrastructures 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) in their study on the relationship between ethnicity and economic development in China

Herbst (2014) claimed that the difficulties that African political leaders meet in governing countries divided into multiple ethnic groups, combined with the separation of ownership and control of land, drove them to develop infrastructure that would allow them to extend their authority as far as possible; see also Herbst (2000). The bulk of rural people in Sub-Saharan Africa are cut off from political power centres due to the inadequate or non-existence of road networks in most nations. According to Herbst (2000), African governments have been expanding their authority by developing highways connecting political centres to formerly unreachable locations.

The investigations of Muller-Crepon et al. (2020) are based on a thesis that states that public officials' access to ethnic homelands is a requirement for political leaders to extend their dominance over remote areas Tollefsen & Buhaug (2015); Raleigh & Hegre (2009); Buhaug & Rod (2006); Herbst (2000); Fearon & Laitin (2003). Certainly, Muller-Crepon et al. (2020) claim that political instability is more prevalent in areas where political power has little control over the population; more specifically, they claim that ethnic groups' homelands, which are difficult to reach from the national capital, are also hotbeds of rebellion and political conflict. The authors utilized a map of African roads' networks to calculate access to ethnic homelands from national capitals and ethnic group inter-connectivity. When the effects of these two variables on the occurrence of political conflicts are estimated, they show that political conflicts occurred in areas where the state had a weak physical presence. To demonstrate the robustness of this finding, they built a theoretical network comprised of hypothetical road segments that ensure the shortest travel time between the busiest ethnic groups; using these hypothetical networks, they build instrumental variables estimating the accessibility of ethnic homelands from political centres as well as the inter-connectivity of ethnic groups if these networks existed. Their findings are robust to the use of these instruments, confirming the dual role of roads' networks in extending political power control and reducing political instability; political authority and stability are strengthened in ethnic groups that are accessible from political capital.

To construct our instrumental variables for the proximity to road networks and its interaction with high-speed internet, we rely on this literature, which highlights political leaders' motivation to build roads in order to expand their capacity to govern. In doing so, we follow a large body of literature that shows that non-economic factors influence African governments' willingness to invest in public infrastructures, as stated in Bonfatti et al. (2019); Jedwab & Storeygard (2021); Burgess et al. (2015); Herbst (2014) have stated. 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 city to the rest of the country Herbst (2014); Muller-Crepon et al. (2020).

Following these assumptions, we use a map of African ethnolinguistic divisions created by American anthropologist in 1959 by Murdock (Murdock (1959)) to identify the ethnic homelands within all countries of our sample. These ethnic homelands are generally regarded as being exogenous. Using the least-cost path algorithm, we build hypothetical roads that connect the centroids of different ethnic homelands to the capital city of each country. We use the assumption that the social planner's goal is to keep the construction costs of these roads as low as possible. The hypothetical and actual roads' network of Ethiopia represented in figure (3-6 and 3-7) show how such a network would evolve to connect ethnic homelands with the capital city if only the accumulated geographical costs of networks' construction between two locations were considered and all other economic factors (including demand-side factors) ignored.

We use the proximity to these hypothetical roads networks to compute instrumental variables for the proximity to the actual road networks, and that of its interaction with the Internet. The validity of these instrumental variables relies on the fact that the hypothetical road networks must be close enough to the actual networks and must be orthogonal to the economic activity and job creation by completely ignoring demand-side factors and local labor-market conditions. In our analysis, the two conditions are most likely to be met. Figure (3-7), for instance, compares the dynamic of the hypothetical road networks to that of the actual road networks in Ethiopia. As illustrated by the map, these two networks are close enough to satisfy the latter conditions. Moreover, the fact that ethnic homelands are exogenously located makes it likely that the hypothetical road network is exogenous to local economic development. In addition, the fact we control for the 20*20 km grid cell fixed effect means that we are comparing the outcomes of individuals and firms that are located in places that are homogeneous with respect to a wide range of geographical, cultural and historical factors (such as the slave trades and the presence of ancient cities).

The Causal Effect of Internet : (DID) Approach

Connectivity to Internet may possibly differ in various aspects which are all correlated to the development and growth process within countries. In fact, the first Internet access in sub-Saharan Africa in the early 2000s was initially transmitted through telephone cables. Indeed, at that time, most if not all of the Internet network was made of terrestrial cables built by telecommunications companies. These terrestrial cables had several limitations likely to increase costs, in particular, the cost of communication, large installation costs of terrestrial cables and long distances traveled by these cables involving a long time for information to be transmitted. There is also the risk of destruction of these cables due to bad weather such as storms, lightening and tornadoes. Clearly, the construction and maintenance of these terrestrial cables were not only expensive but strongly correlated with the evolution of economic activity, since this essentially monopolistic market only included for the most part, large private or governmental companies capable of bearing such costs. Another essential limit to this type of Internet connection was the transmission or broadcasting distance, which was quite short.

The arrival of submarine cables in sub-Saharan Africa beginning in the year 2009 appears as a solution to many of the limitations set out in the previous lines, because not only it lowers installation and maintenance costs, but the information transport times are also shorter, as well. It is for these purposes that submarine Internet arrived in sub-Saharan Africa, mainly outside the major economic centers whose cities are connected to terrestrial cables networks. The arrival of the submarine Internet has brought higher signal transmission speed and signal emission distances, allowing for better traffic of information between Africa and the rest of the world. In figure (3-4), we can appreciate how much the terrestrial network has evolved over time, especially since the arrival of high-speed Internet.

We exploit the exogenous nature of the different arrival dates of submarine cables on the coast of African countries south of the Sahara to establish the causal effect of Internet on our labor market outcomes. Indeed, each country is treated on a very specific date; the date on which the very first submarine cable arrived at the coast of sub-Saharan Africa and connected to the land network of the country. This makes it possible to hypothesize an absence of spillovers between countries. We thus define two groups, the group of treated individuals, made up of those who therefore have access to high-speed Internet (submarine Internet); this is the group of individuals who are located within a radius of 2 kilometers from the nearest terrestrial network after the arrival of the submarine cable. The group of non-treated individuals is composed either of people (or firms) being located more than two kilometers from the nearest terrestrial optical fiber, or those connected (or not) to the Internet network before the arrival of the submarine cables.

Causal identification relies on the assumption that in the absence of submarine Internet, the characteristics of the treated and the control groups would have similar trends. However, it is not possible to prove it empirically. Instead, we show that the two groups present similar trends in their characteristics before the arrival of submarine Internet. We display the time trends of our main outcome variables in figures (3-9)-(3-10). We also compare the characteristics of the treated and the control groups before the arrival of submarine Internet (table (3-5)).

3.4 Findings

3.4.1 First stage results

Whether it is the instruments for the road or that of its interaction with the Internet treatment variable, we consider these instruments valid for at least two reasons. First, when we compare our theoretical network in the figure (3-6) and the real network, one can note a great similarity in its structure. The results of the estimation of equation (3.1), effectively show that there is a positive and very significant relationship between the distance to the actual road network and the distance to the hypothetical road network (table (3-6)). Just as it shows a statistically non-zero relationship between the distance to the road and the topography of the land. These results are consistent across the three databases we use in this work. Similarly, we find a significantly positive correlation between the real interaction term and the conjectured one, as well as a statistically negative correlation between the interaction term and the lightning intensity.

For our three samples, a strong and significant correlation is demonstrated between the instruments used and the non-exogenous variables of our model. Especially with the statistical tests presented which reveal that the chosen instruments are strong enough, in light of the Cragg-Donald Wald or Anderson-Rubin Wald tests. Certainly, they all allow us to reject the null hypothesis of weak identification. Also, the Kleibergen-Paap test allows us to reject the null hypothesis of under-identification. Our first stage is, therefore, robust across the three samples used, with the theoretical model well fitted to the data.

3.4.2 Average Joint Effect of Roads and Internet

This subsection presents the average joint effect of Submarine internet and road on employment and occupation types, by estimating equation (3.2).

Employment

Table (3-7) displays the coefficients resulting from the estimation of equation (3.2). In column (1) and (2), using the DHS sample, we regress two measures of the odds of an individual to offer their labour force on the infrastructure variables and on all our controls including year, country and grid cell fixed effects. In the third column, we use the LSMS data to compute the effect of infrastructures on the number of hours worked, using the same identification strategy as in the first two columns.[8]. The last 3 columns display the estimations computed from the WBES, with the dependent variables being the number of hours of operation per

[8]Information on the number of hours worked is not collected in the DHS

week in last fiscal year, the number of permanent full-time employees at the end of the fiscal year lagged three years, and the number of full time employees in the last fiscal year, respectively.

Given a specific distance to the closest main road, an individual connected to submarine internet is 8.3 percentage points more likely to have had an employment during the last twelve months preceding the interview. However, when that individual gets one kilometer closer to the main road, his odds to have been working during the last year increases by 10 percentage points in comparison to someone who has also gotten one kilometer closer to a main road, but who is not connected to high-speed internet (DHS results column (1)). Likewise, when individuals with no access to internet are closer to a major road by one kilometer, their likelihood to have offered their labour force in the year preceding the survey increases by 0.8 percentage points. However, if the accessibility to submarine internet is combined with the increase in the proximity to the road, an average individual's likelihood to have worked in the preceding year increases by 2.5 percentage points. Hence, internet and road complement each other to boost employment.

In column (2), for a fixed distance to the main road, individuals connected to submarine internet are 5.7 percent more likely to be working, compared to those not connected. Similarly, in absence of internet access, the isolated effect of marginal increase in the proximity to a main road raises the odds to be working by 0.8 percentage points. Therefore, being connected to submarine internet when getting closer to a major road by a kilometer increases the likelihood of being employed by 1.2 percentage points; which again clearly shows the complementary between internet and road. These results suggest that road and internet infrastructures act as complements to increase employment opportunities by an average of 1.2 percentage points.

Figure (3-11) clearly illustrates the gap between the effect of the road on employment for people connected to the submarine internet compared to those who are not connected. At equal proximity to the nearest major road, the group not connected to submarine internet has about 0.8 percentage points probability of being employed at the time of the survey, while the connected group has about 2.1 percentage points probability of being employed, which represents a gap of about 1.3 percentage points between the two groups. This result is consistent with the coefficient of the interaction between the road and internet, and validates the hypothesis according to which, gaining access to the Internet reinforces the positive effect of roads on employment.

In column (3), we find that for a given distance to the closest main road, individuals connected to fast internet work approximately 20 fewer minutes a week than those not connected. This negative sign can be explained by the increase in the productivity generated by the internet Hjort & Poulsen (2019b); Rollo & Paunov (2015). However, when the distance to the closest road decreases by a kilometer, the number of hours of work per week for individuals connected to submarine internet increased to 3.49 hours. As for the isolated effect of the road, the number of hours worked per week increases by about 57 minutes as individuals get one kilometer closer to the main road. Indeed, with a better proximity to a main road, the travel time to work certainly decreases, which could explain this positive effect of road on the number of hours worked per week. The addition of the accessibility to high-speed internet to a better accessibility to the road network gives an extra boost to the number of hours worked by 4 hours more per week. So, roads and internet complement each other to give at least 3.5 extra hours of work.

From the WBES, we find that in absence of internet, an increase in the proximity to the closest central road increases a firm's operation hours by three hours and 46 minutes. For an equal change of one kilometer in the closeness to the nearby main road, connected firms operate approximately 6 minutes longer than non-connected ones. Similarly, holding fixed the distance to the closest central road, firms connected to internet operate approximately 35 minutes longer compare to non-connected ones. Moreover, there is a gap of 6 minutes in the amount of time of operation between a connected firm that has gotten one kilometer closer to the main road and its counterpart who is connected and is located within an unchanged distance to the main road (column (4)). Therefore, getting for example ten kilometers closer to the nearest main road when gaining access to the internet increases the number of weekly operation hours of the firm by 60 minutes. This means that the road and the internet complement each other to increase firms' labor demand.

This result is consistent with those listed in columns (5) and (6) related to the number of permanent employees. We find that the road and the internet accessibility act as complements to generate a further increase in the number of permanent workers in firms. In column (5), internet-connected firms that move 10 kilometers closer to the nearest main road have 0.27 more permanent workers at the end of the fiscal year three years ago compared to firms that have experienced an improvement in their accessibility to infrastructures only for the road or the Internet in isolation. The result in column (6) reinforces the latter because in the last fiscal year the gap in terms of the number of permanent employees for the 2 groups of firms is 0.72 more workers for the first group of firms.

Those results suggest a positive and highly complementary effect of road and accessibility to internet in the enhancement of employment. In other words, supplying both forms of infrastructures at the same time creates considerable job opportunities. According to the estimates in table (3-7), being closer to a major road improves the effect of being connected to submarine internet on the likelihood of employment, with a statistical significance of 1 percent. When road and internet access are provided at the same time, they support one other, resulting in job creation. Certainly, they are highly complementary in the sense that providing one greatly boosts the marginal effect of the other. This complementary holds true across all of the tree outcome variables we have looked at so far, as well as between the data sources. The effects of infrastructure on the types of employment and occupational sectors are our next investigation.

Skilled versus unskilled employment and occupational sectors

In table (3-8), we present the results of the investigation of both the distinct and joint effects of roads and internet access on employment by level of qualification. Connectivity to high speed internet, coupled with a better proximity to major roads increases employment in professions requiring high qualification and diminishes employment in unskilled occupations; this favors transition from low-skilled employment to higher skilled professions.

Our findings indicate that the marginal effect of supplying high-speed internet in locations within a fixed distance to a major road gives individuals lowers the probability of being in a low skilled profession by 13.1 percentage points. Moreover, providing high-speed internet to individuals who get one kilometer closer to a major road leads to an additional 2 percentage points decrease in their likelihood to have low-skilled employment. Even though the isolated effect of the road tends to significantly increase low-skilled employment, getting one kilometer closer to the closest main road, when connected to submarine internet decreases the likelihood to have a low-skilled job by 1.4 percentage points. In summary, road proximity and internet accessibility complement each other to reduce the low-skilled employment by 1.94 percentage points (column (1)).

In the third and fourth columns of the same table, we look at the same effect on skilled and high-skilled employment. The joint effect of the two types of infrastructure is positive and highly significant, showing that they play complementary roles in the creation of skilled employment. The marginal impact of getting one kilometer closer to the nearest main road increases for individuals in locations connected to submarine internet. Similarly, the marginal impact of internet accessibility increases as the distance to the closest main road decreases. Indeed, being one kilometer closer to the main road increases skilled employment by 2.2 percentage points for individuals in connected locations. We obtain similar results for workers who offer services; offering services might require a minimum level of qualification. When supplied simultaneously, roads and internet complement each other to give an extra increase of approximately 3 percentage points to services employment (column (3)). This result is not only consistent with that of the skilled employment in column (3) but also, is very significant at the one percent level.

Figure (3-12) is consistent with this further finding as it shows that for an equal change in the proximity to the road, the likelihood o being a skilled worker is about 0.01 percentage points in non-connected areas, while it is around 1.86 percentage points in connected locations, which is equivalent to a gap of 1.85 percentage points between the two groups. Clearly, road and internet complement each other to give an extra boost to the increase in the demand for skilled labour force by roughly 2 percentage points.

As for highly skilled jobs, we find a statistically null effect of the road, taken in isolation. However, for localities connected to fast internet, individuals who are one kilometer closer to the main road are 1.6 percentage points more likely to be employed in very highly skilled professions (column (4)). It is very interesting to note that, while road and Internet accessibility complement each other to further reduce the number of unskilled jobs by about two percentage points, they do the same to increase the creation of skilled jobs by an average of two percentage points. This makes it possible to conclude unequivocally that the development of road and digital communication infrastructures act as complements to allow migration of employment from unskilled sectors of activity to those requiring a certain level of qualification.

These results are consistent with those obtained from the WBES relating to the number of full-time employees with a high level of qualification. Road and internet play a complementary role in increasing the number of qualified full-time workers in companies. In column (5), internet-connected firms that moved 10 kilometers closer to the nearest main road had 0.372 more high-skilled full-time employees at the end of the last fiscal year compared to firms that have experienced an improvement in their accessibility to only one of the two infrastructures. The result of column (6) reinforces the hypothesis that the combination of the two infrastructures additionally improves employment, even for the suppliers of unqualified labor. Indeed, we show empirically that firms in the first group have 0.621 more unskilled workers than those in the first group. The combination of these two results (columns (5) and (6)) simply means that the firms in the first group have a statistically higher labor demand, both for skilled and low-skilled labour force, compared to the second group of firms who obviously demand less of both categories of the labour force. These results obtained on the WBES sample not only support the hypothesis that road and the Internet together yield an additional enhancement to job creation, but they also do not contradict the result of labor migration from unskilled jobs to higher qualification employment. Unfortunately, because of the significant difference in the number of responses in columns (5) and (6), it is difficult to determine whether, in the WBES respondent firms, the joint effect of our two infrastructures is greater on the skilled or unskilled employment.

Agricultural Employment

Estimating equation (3.2), with agricultural employment as the outcome variables shows that enhanced access to roads and the internet networks increases the prospect of employment (Table (3-7)) and stimulates labor reallocation, with labour force shifting from farming and other low-skilled occupations to skilled professions (table (3-8) and table (3-9)).

Table (3-9) displays the effect of our two types of infrastructure on agricultural employment. We distinguish between self-employed farmers and farmers in general. Overall, we find that in locations connected to submarine internet, when situated within a fixed distance to the closest main road, individuals are 14.7 percent less likely to be self-employed farmers, in comparison to their counterparts located in connected localities and within an equal distance to the closest main road. Moreover, when getting one kilometer closer to the main road, individuals in connected localities lose an extra 2.8 percentage points probability of being self-employed farmers. Even though a better proximity to road in absence of internet connectivity increases the odds to be a self-employed farmer, when supplied simultaneously, both infrastructures reduce the prospect of individuals operating as autonomous farmers. We obtain similar results for agricultural employment overall. Indeed, the marginal effect of an increase in the proximity to the nearest major road for locations connected to submarine internet, decreases by 2.6 percentage points, the probability of individuals working in the agricultural sector. Figure (3-13) confirms this result by showing a difference greater than 2 percentage points in the effect of the road on employment in the agricultural sector, between areas connected and those not connected to submarine Internet.

Furthermore, given the critical role that agricultural activity plays in less urbanized localities, we are now paying more attention to the causal effects of road and internet supplied simultaneously on agricultural employment. We investigate the isolated and joint effects of the better provision of these two infrastructures on agricultural activities for rural populations. The coefficients of the regressions are presented in columns (3) and (4). They exhibit results that are consistent with those in the first two columns. Better accessibility to roads and the internet reduces agricultural employment in rural areas significantly. Being one kilometer closer to the nearest main road, in a location with fast internet access, diminishes the probability of working in agriculture by 2.97 percentage points (Column 4). Likewise, access to submarine internet in a location along a road, reduces agricultural employment by nearly 1.1 percentage points. Hence, the marginal impact of being one kilometer closer to the nearest main road is 1.7 percentage points smaller for individuals in internet-connected locations.

This result is consistent with that shown in Figure (3-14) where we can clearly see that there is no statistically significant difference between connected and unconnected regions in the effect of the road on urban agricultural employment. This is obviously not the case in rural areas.

Referring to tables (3-7) and (3-8), gaining internet access in conjunction with improved road access has a comparable impact on the landscape of employment at the regional level. Internet and road accessibility jointly create new job opportunities (table (3-7)), mostly in the service sector (column (2) of table (3-8)). Individuals with specific skills may also be able to transition from low-skilled occupations to positions necessitating higher qualifications.

Finally, the two types of infrastructures have greater effects in combination, on employment and job creation overall, but also favor the migration of the labour force.

3.4.3 Heterogeneous Effects of Roads and Internet on Employment

This section investigates heterogeneities in the joint effects of being closer to roadways and having access to internet. Our findings reveal that the outcomes differ depending on the level of a country's economic development and by individual characteristics including gender, age, and place of residence. This also provides some insights regarding the channels through which infrastructure investment influences employment. Recall that the effects of road and Internet are heterogeneous across types of jobs and qualification level. Therefore, if individual traits influence the prospect of employment in specific industries, then, infrastructure investments are prone to have various outcomes based on specific individual attributes.

Economic Development

We look at how the impacts of roads and submarine internet on employment vary across levels of economic development. Different professional choices amongst countries with different levels of economic development could theoretically cause mixed outcomes. Agriculture, for example, may be more prevalent in developing economies, while workers offering services may be more predominant in developed countries. We categorize countries' development level based on the 2019 GDP classification summarized in table (3-1).

The results of the regression of equation (3.2), distinguishing between more and less developed countries are reported in columns 3, 4, 7, 8 11, and 12 of table (3-10). The effects of road expansion on employment differ by level of economic development. In areas with access to the internet, increasing proximity to the road network by one kilometer does not statistically increase the probability to have offered their labour force during the twelve months preceding the interview in more developed. In less-developed countries, the corresponding effect is 1.2 percentage points, and it is highly statistically significant (column

(4)). We obtain similar findings regarding the probability of being employed in columns (7) and (8). In areas where the road network gets ten kilometers closer to the community, being connected to submarine internet increases the probability of working by 9 percentage points in less-developed countries, compared to a non-statistically significant effect in more-developed countries.

Clearly, there is a positive and highly significant complementary effect of road and internet on employment in less-developed countries, and that effect is not present in developed economies. In locations that are far enough from road without any internet access, gaining access to both infrastructures has a positive effect, but this effect is strongly statistically significant only among less-developed countries. These findings imply that the complementary role of roads and Internet is stronger in economies with low level of development.

In columns (11) and (12), the findings are similar regarding the number of hours worked during the last seven days. There are higher and significant effect across low-developed countries, with combined the marginal effect of the road and Internet being twelve additional hours of work per week in less developed economies, reflecting the low employment rate in less developed countries prior to fast Internet arrival.

In summary, the isolated effect of accessibility to the submarine Internet is non-zero only in poorly developed countries, while the isolated effect of the road is statically non-zero in both types of regions, but the latter is higher in developed countries. However, when better proximity to the road is added to accessibility to high-speed Internet, the joint effect of both infrastructures increases employment only in poor countries.

Rural/Urban Differentials

We look at how the effects of extending roads and internet access on employment differ in urban and rural areas. Table (3-10) shows the results in columns 1, 2, 5, 6, 9, and 10. In both urban and rural areas, road expansion has resulted in net employment growth. However, the marginal employment effect of roads is greater in urban locations. According to the estimates, decreasing the distance to the road network by one kilometer raises employment by 9.7 percentage points in urban regions and by 8.22 percentage points in rural areas among communities connected to the submarine internet. Similarly, in communities with a fixed distance to the nearest major road, gaining internet access increases the probability of working by 30 percentage points in urban areas and by a zero percentage points in rural areas. Furthermore, only in metropolitan regions is the combined marginal effect of road and internet access statistically non-zero. The justification for this could simply be that rural areas need more roads than they do need internet connectivity. Indeed, the low level of education in rural areas may simply undermine the effect of the Internet; so only roads development allows job creation in rural areas.

We could not find any significant joint differential effects in the number of hours works between urban and rural areas.

We also analyze the differential effects of roads and internet on low-skilled and high-skilled employment in urban and rural areas. The results are reported in table (3-15). The findings show that low-skilled and high-skilled employment respond similarly to the expansion of infrastructure. Under the joint effect of both infrastructures, high-skilled employment increases in rural areas (column 3), but have a null impact in urban areas (column 2). Similarly, low-skilled employment tends to decrease in rural areas, with a statistically null effect in urban areas (table (3-17)). In urban areas connected to submarine internet, getting one kilometer closer to the road networks increases the likelihood of working in high-skilled job by 8.1 percentage points (column 2 of table (3-15)). In rural connected areas, getting one kilometer closer to the road networks increases the likelihood of working in a high-skilled occupation by 6.9 percentage points (column 4). While the joint effect of both infrastructures in positive in rural areas, it is nil in urban areas. Therefore, roads and internet complement each other to bring an extra increase in high skilled labour demand only in rural areas.

Similarly, in rural connected areas, getting one kilometer closer to the nearest main road decreases low-skilled employment by 11.34 percentage points, while in urban connected areas, getting one kilometer closer to the nearest main road increases low qualified employment by 0.3 percentage points. In fact, the isolated effect of the road is negative in urban areas, as well as in rural areas. However, the addition of Internet to the road is beneficial in that it decreases low-qualified occupations in rural areas; in these areas, gaining internet access marginally decreases the likelihood of working in a low-skilled occupation by 1.1 percentage points, but the effect is not statistically significant in urban areas (column 2 of table (3-17)).

In the services sector, we can observe that the two infrastructures together complement each other to increase employment (table (3-20)), but with a statistically higher effect in urban areas. While the road and the internet complement each other to bring an additional boost to the demand for labor in the services sector of about 6.3 percentage points (column 3) in urban areas, in rural areas the extra boost is only 1.2 percentage points (column 3).

To summarize, in rural areas, investing in both road and internet infrastructures promotes job migration from low-skilled to high-skilled occupations. In urban areas, bringing the road network closer to communities and simultaneously investing in the development of internet infrastructures essentially promotes job creation in the services sector.

Gender

In this part of the analysis, we look at the differential effects of the simultaneous improvement in road and internet access on employment across genders. Table (3-11) displays the results of regressions estimated for each gender group. We find that there is not a significant heterogeneity of the combined effect of roads and internet on employment across gender. Indeed, both infrastructures act as complements to significantly increase employment for both men and women. But the effect is higher among females. In absence of internet access, getting ten kilometers closer to the main road increases men's probability of having worked during the last twelve months preceding the survey by 4.1 percentage points against 9 percentage points for women. However, gaining access to Internet with those within a ten kilometer proximity to the road, allows male odds of employment to increase to 15.9 versus 28.29 for females. So, the connectivity to internet causes an additional 11.8 percentage points to have worked in the past year for men, and 19.3 for women (columns 1 and 2). In column 3, only the isolated effect of road network improves employment prospects for men, while the combination of the two infrastructures is more beneficial for female labor force providers (column 4).

As for the number of hours worked during the last seven days, we find that Internet access, holding fixed the distance to the main road, tends to reduce the number of hours worked by approximately 68 minutes for men against 27 minutes for women. However, moving toward the road by one kilometer, the number of hours worked during the last 7 days increases to 21 minutes more for men and 61 minutes more for women.

Therefore, the way in which the combination of these two infrastructures affects employment does not differ qualitatively by gender. According to the aforementioned findings, it differs only in their magnitude. We have higher impact among women.

Columns 4 and 5 of table (3-15) present results consistent with the latter. Internet and road complement each other to allow an extra increase in the demand of high skilled labour force for both men and women. But the effects are greater and more significant among women. However, they complement each other to decrease low-skilled employment, with higher significance and magnitude among men. With connectivity access, getting one kilometer closer to the main road decreases the probability of having low-skilled employment by 26.1 percentage points for men versus 10.1 percentage points for women (table (3-17)).

The findings imply that employment gains from investment in road and internet in sub-Saharan Africa are more beneficial for women than for men. However, for some employment outcomes, the effects are larger among men. For instance, the joint effect of these two infrastructures favors the sectoral migration of the labour force from the low-skilled sector to the high-skilled sector more among males than females. Also, in the WBES, the joint effect of these infrastructures is statistically null for women. In fact, in column (8) of table (3-11), we find that taken in isolation, Internet has a greater impact on the demand for the male labor force, while roads have a greater impact on women's labour supply. However, their joint effect significantly increases the demand for labor only for men.

Age Group

It is possible that the joint investment in the development of road networks and internet infrastructures affects individuals differently depending on the age group to which they belong. In order to verify if there is any differential effect, we divide our sample into three distinct age groups: 20-34, 35-44 and 45-64 years old. The results of the regressions are reported in table (3-13). For the outcome variable indicating if individuals have worked during the year preceding the survey, across all three samples, the results show higher net employment gains for the 20-34 age group than for the two other age groups. Indeed, for this outcome variable, individuals aged between 20 and 34 years old residing in locations connected to submarine internet, when they get one kilometer closer to the nearest major road, are 2 percentage points more likely to have worked during the year preceding the interview, compared to their counterparts located to an equal distance to the closest main road, but who have no internet access (column 1). In the 35-44 age group, the joint effect of both infrastructures gives 1.1 percentage points more odds of employment and 1.4 in the 45-64 age group (columns 2 and 3).

We find great similarities between these effects on all our outcome variables capturing employment. As with the first outcome, the effect of the two infrastructures on whether an individual is currently employed is greater and more significant in the 20-34 age group, whereas the combination of the two infrastructures does not increase employment significantly for the other age groups. Only isolated effects are significant (columns 7 and 8). As for the number of hours worked in the last seven days, the joint effect of the two infrastructures increases the number of hours worked much more in the the younger age groups. Even though our two infrastructures do not significantly affect the number of hours worked among people aged 45-64 (column 11), we nevertheless find a positive coefficient.

In tables (3-15), (3-16) and (3-17), we are interested in how the effects of road and Internet on low-skilled and high-skilled employment vary by age. We find that the two infrastructures jointly exert a positive effect on high-skilled employment and a negative effect on low-skilled employment. However, the effects are higher and more significant for the youngest. In other words, the road network and Internet act as complements to favor a shift of employment from low-skilled to higher qualified jobs, but with a greater magnitude among the youngest.

In summary, these results imply that younger labour force suppliers are more likely to benefit from the simultaneous expansion of roads and ICT infrastructures, presumably as workers in this age group are less refractory to technological change.

Education

Lastly, we investigate how the effects of road and Internet differ by level of educational attainment. For this, we create four educational groups in our sample: (1) individuals who have no education at all; (2) individuals who attended only the primary school; (3) individuals who did not go further than the primary school (including non-educated individuals); and (4) individuals who have at least a secondary education level. From table (3-12), it appears that the prospect of employment tends to be higher among the less educated individuals under the joint influence of road and Internet (see the coefficients on the interaction term in columns (1) till (8)). Indeed, it can be seen that the simultaneous investment in both types of infrastructures increases employment opportunities, much more for the least educated group. This could be explained by the fact that the marginal benefits of investment in these two infrastructures on knowledge acquisition are likely to be greater for individuals in need of education. This presumption is precisely confirmed in tables (3-15), (3-16) and (3-17), where we can see that the joint effect of the two infrastructures further increases the probability of working in a high-skilled profession by 2.18 percentage points for the least educated individuals, whereas the additional increase is only 0.7 points for those with primary education, and 1.4 for those with at least a secondary education. Similarly, in localities connected to submarine internet, getting one kilometer closer to the main road decreases the probability of working in a low-skilled job more in the group of less-educated workers, with a nil effect for the most educated people. However, their effect increases the number of hours worked among all education levels.

Therefore, with the exception of their effect on the number of hours worked, the joint effect of road and Internet expansion is more beneficial for less educated individuals, thus favoring their migration to more skilled jobs.

3.5 Joint Effect of Road and Internet on Firms' Operations and Productivity

We use exactly four variables to assess the joint effect of road and internet on the operations and productivity of firms that responded to the WBES : (1) a dichotomous variable taking the value one if the firm asserted that access to telecommunication infrastructures is a brake on its operations, and zero otherwise; (2) percentage of annual losses in annual sales volume;

(3) the percentage of the firm owned by the government; and finally (4) an indicator for whether the company has secured at least one contract with the government during the last fiscal year.

In column (1) of table (3-14), we find that, for a firm connected to the internet, holding fixed the distance to the road network, the probability that telecommunications be an obstacle to the smooth running of its operations is 6.7 percent less compared to a firm that is not connected to the internet. Therefore, taken in isolation, Internet access could improve the productivity of firms by promoting the smooth running of their operations. A reduction in the distance to the road of 10 kilometers for internet-connected firms gives them an additional reduction in the obstacles to their productivity of 3.6 percentage points. Conversely, in absence of internet connectivity, a reduction in the distance to the nearest main road by 10 kilometers would increase telecommunication barriers by 36.2 percentage points. However, by gaining internet access, these firms reduce this barrier to production by 3.6 percentage points. Thus, road and Internet play a complementary role in reducing barriers to productivity.

In column (2), we examine the effect of infrastructure on annual losses as a percentage of total sales in the last fiscal year. For a fixed distance to the nearest main road, a firm not connected to the internet has 2.3 percentage points more losses as a percentage of total sales compared to a firm in the same situation, but which is connected to internet. When the distance to the nearest road reduces by 10 kilometers, these losses are also reduced by an additional 1.5 points more for connected firms. Similarly, for unconnected firms, those whose distance to the adjacent main road is reduced by 10 kilometers experience a reduction in their losses by 7.1 percentage points, while their peers who are connected to internet experience a reduction in their losses by 8.6 percentage points in total. Thus, when a firm is closer to the nearest main road by 10 kilometers, internet connection decreases its annual loss by 1.5 percentage points.

Columns (3) and (4) focus on the joint effect of road and internet on private and governmental entrepreneurial initiatives. Column (3) shows that for firms located at a fixed distance to the nearest main road, those connected to Internet have 1.3 percent more of their company owned by the government than those who are not connected. For connected firms, getting closer to the road by 10 kilometers increases government shares by 0.85 percent. However, for companies that are not connected, being 10 kilometers closer to the nearby main road reduces the shares held by the government by around 3 percent, but the addition of internet accessibility lowers these shares from 3 percent to 2.1 percent. Therefore, while investment in telecommunication infrastructures favors government entrepreneurial initiatives, investment in the development of roads networks encourages private entrepreneurial initiatives.

The results in column (4) present the effects of the two infrastructures on the propensity of firms to obtain a contract with the government. Internet connection increases the probability of securing a contract with the government of 17.2 percentage points. Moreover, among internet-connected businesses, those that are 10 kilometers closer to the nearest main road are 1.1 percentage points more likely to obtain a contract with the government (18.3 percentage points for those connected firms compared to 17.2 percentage points for non-connected firms). Although the isolated effect of the road is statistically null, its negative coefficient is consistent with the result of column (3) expressing a negative but not significant effect of the road on the propensity of governments to operate in collaboration with the private sector.

To conclude this section, the two infrastructures play complementary roles in enhancing productivity, in particular by reducing obstacles to the smooth running of firms' operations. But, they act as substitutes on private and governmental entrepreneurial initiatives.

3.6 Conclusion

We show that road and Internet complement each other to boost employment through their joint effect on private and public entrepreneurship initiates, and induce a shift of labour force from low-skilled to higher-skilled occupations.

We estimate the average and heterogeneous impacts of road construction and Internet accessibility on employment creation in Sub-Saharan Africa. We look at the effects of increasing access to road and Internet alone and in combination. We discover that both forms of investments have considerable and positive complementary effects. Those positive effects vary by country, gender, age group, and education level. In more advanced economies, the benefits of investing jointly in roads and internet tend to be greater. Within countries, the advantages of expanding both infrastructures differ between urban and rural locations. Roads benefit rural areas more than urban areas in terms of job creation. Furthermore, due to the greater influence of high-speed Internet in urban occupations, the complementarity between the two infrastructures is significantly stronger in these areas.

The simultaneous expansion of road and Internet leads to a decrease in low-skilled employment, particularly agricultural employment, as individuals migrate to higher-skilled jobs. We also find that the combination of both infrastructures increases firms' demands for skilled labour force, as well as unskilled labour force in the LSMS sample, where demand is greater for skilled labor. Also, the simultaneous provision of road and Internet connection increases firms' productivity. Indeed, we find that both infrastructures complement each other to reduce the annual burden of private sector enterprises.

Although both men and women profit from improved road and Internet connectivity, women benefit more. Moreover, younger people benefit more from improved infrastructure access in terms of job opportunities. Our findings suggest that the joint expansion of roads and internet is a major factor in Sub-Saharan Africa's structural development and transformation. Therefore, development policy measures that target job creation must integrate the simultaneous development of these infrastructures to be optimal.

3.7 Appendix

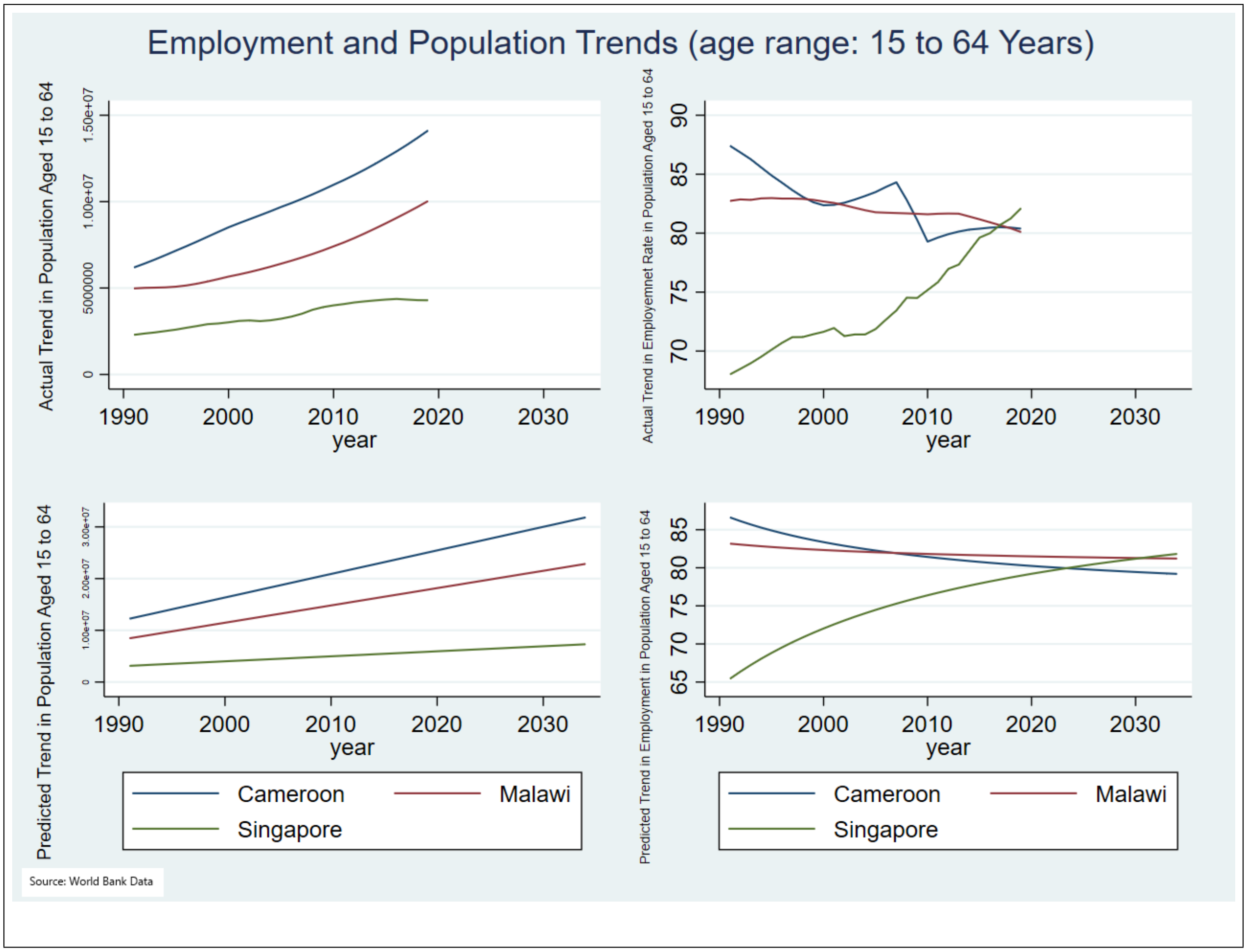


Figure 3-1: Trends in Employment and Population for group aged 15 to 64 years.

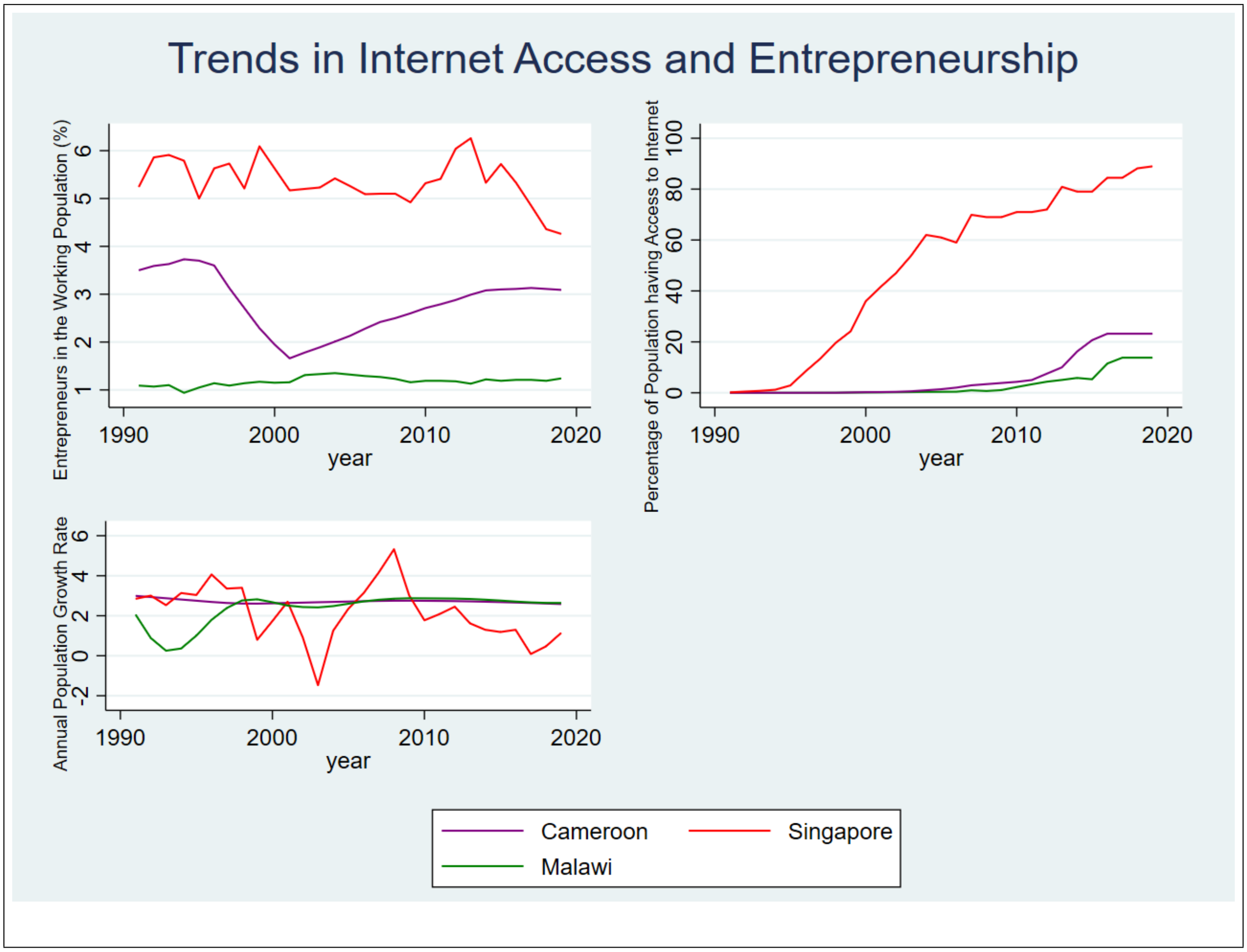


Figure 3-2: Trend in Internet Access and Entrepreneurship .

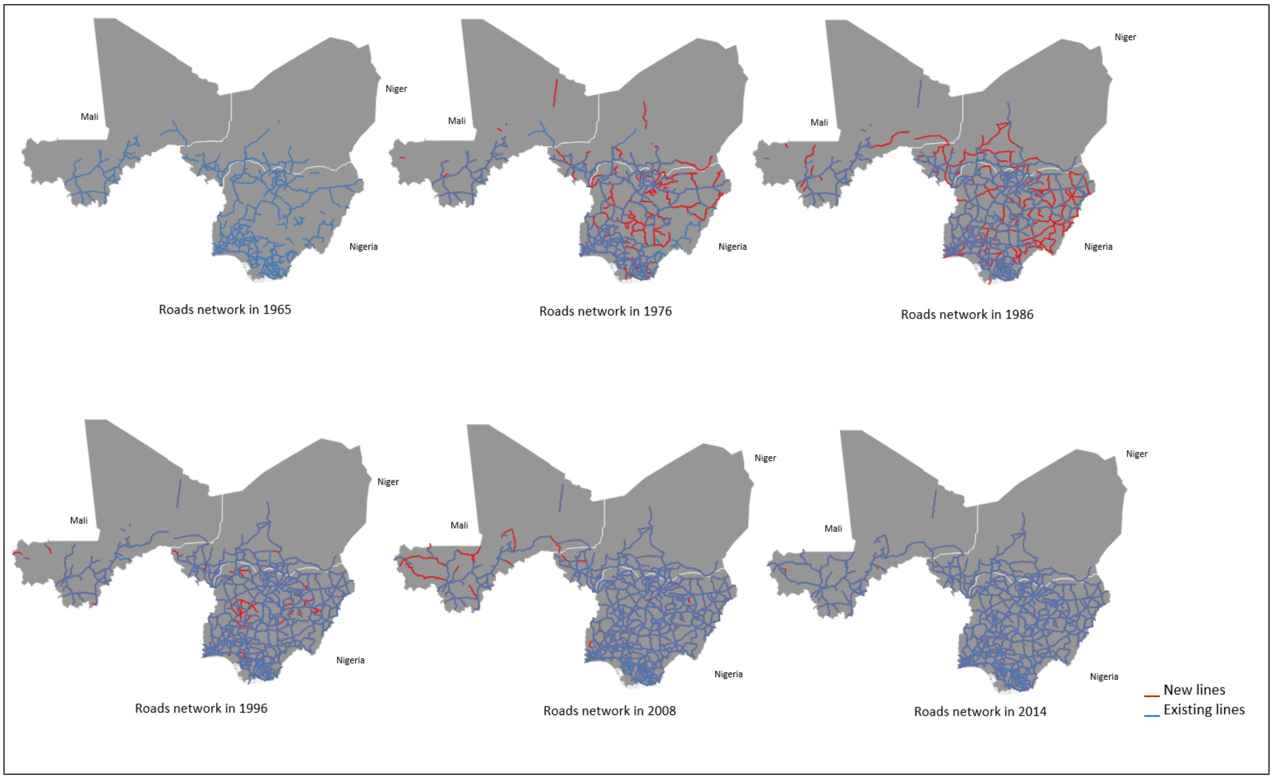


Figure 3-3: Dynamic in North-West Africa's roads network between 1965 and 2014.

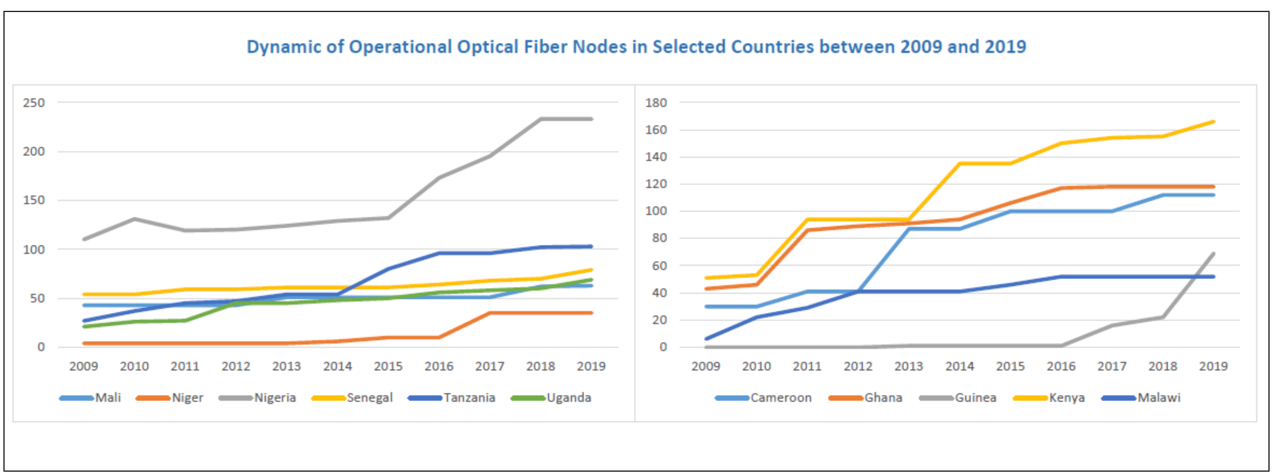


Figure 3-4: Dynamic of fiber optic network between 2009 and 2019.

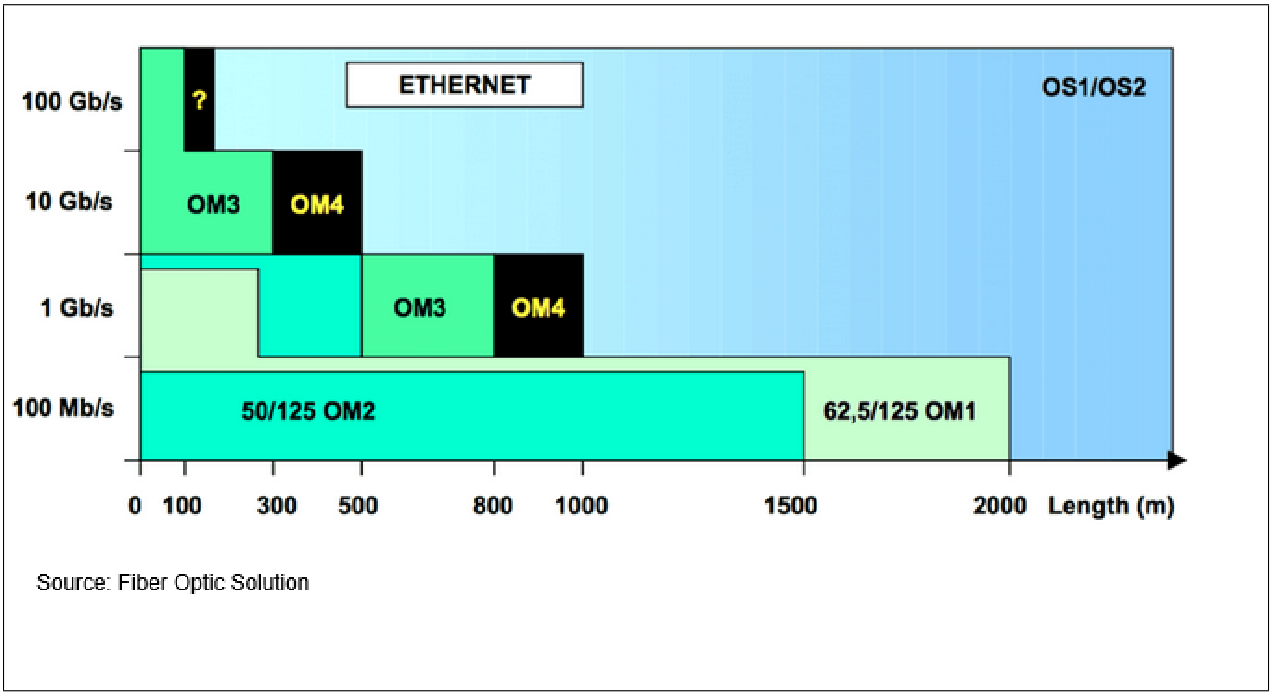


Figure 3-5: Fiber Optic Cable's Signal Emission Distance.

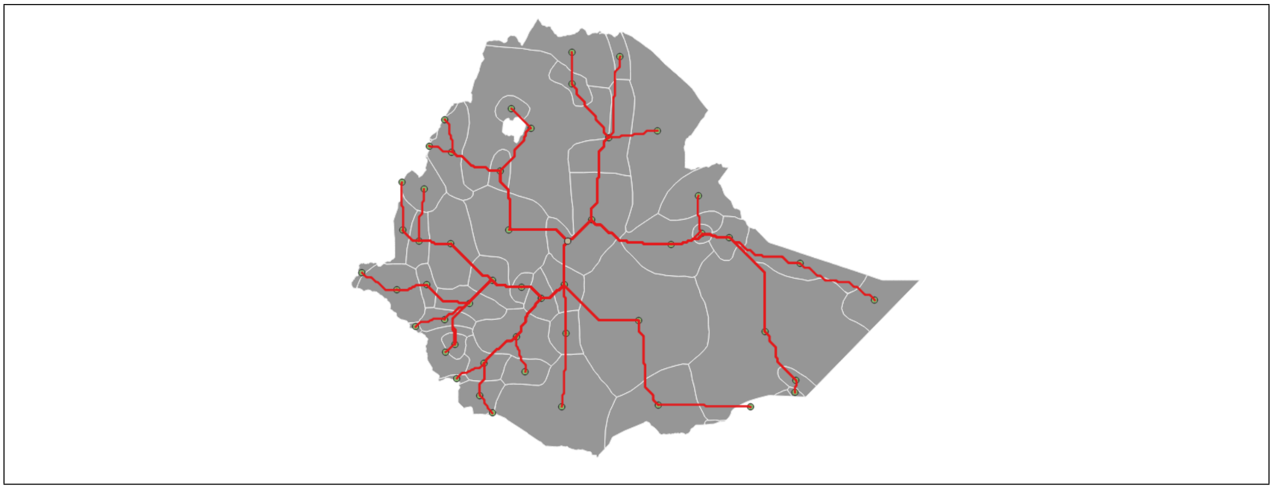


Figure 3-6: Least Cost Path based on Murdock 1959's Map.

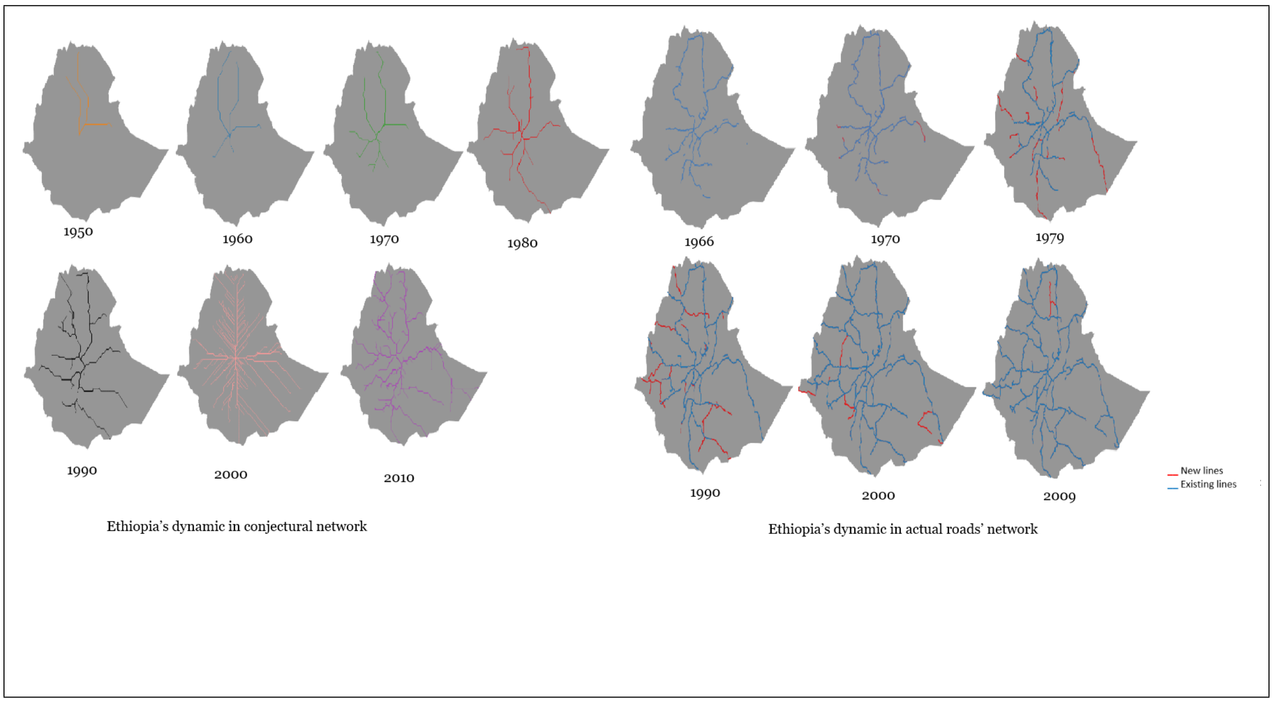


Figure 3-7: Conjectural vs Actual roads' network.

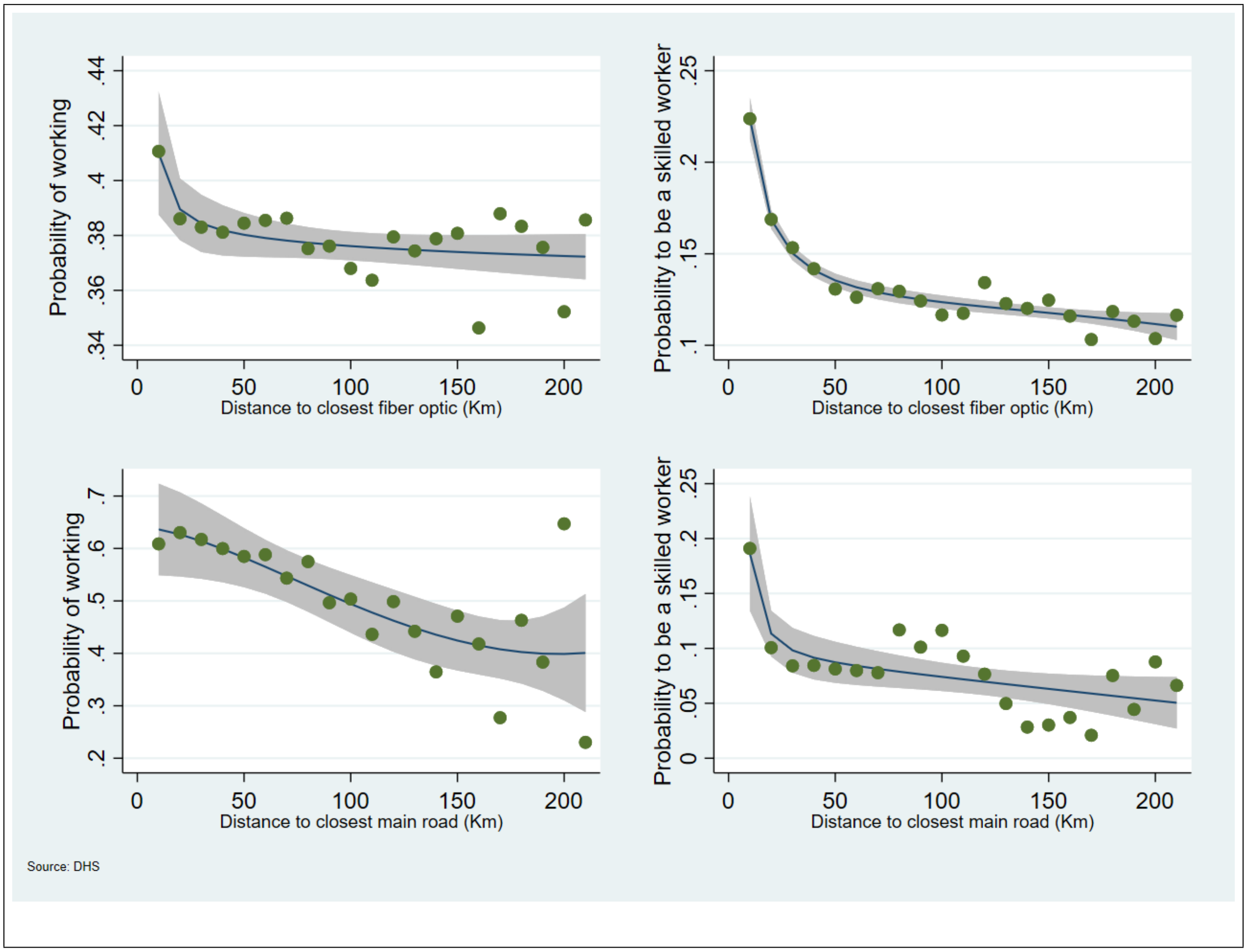


Figure 3-8: Employment Pattern and Infrastructures.

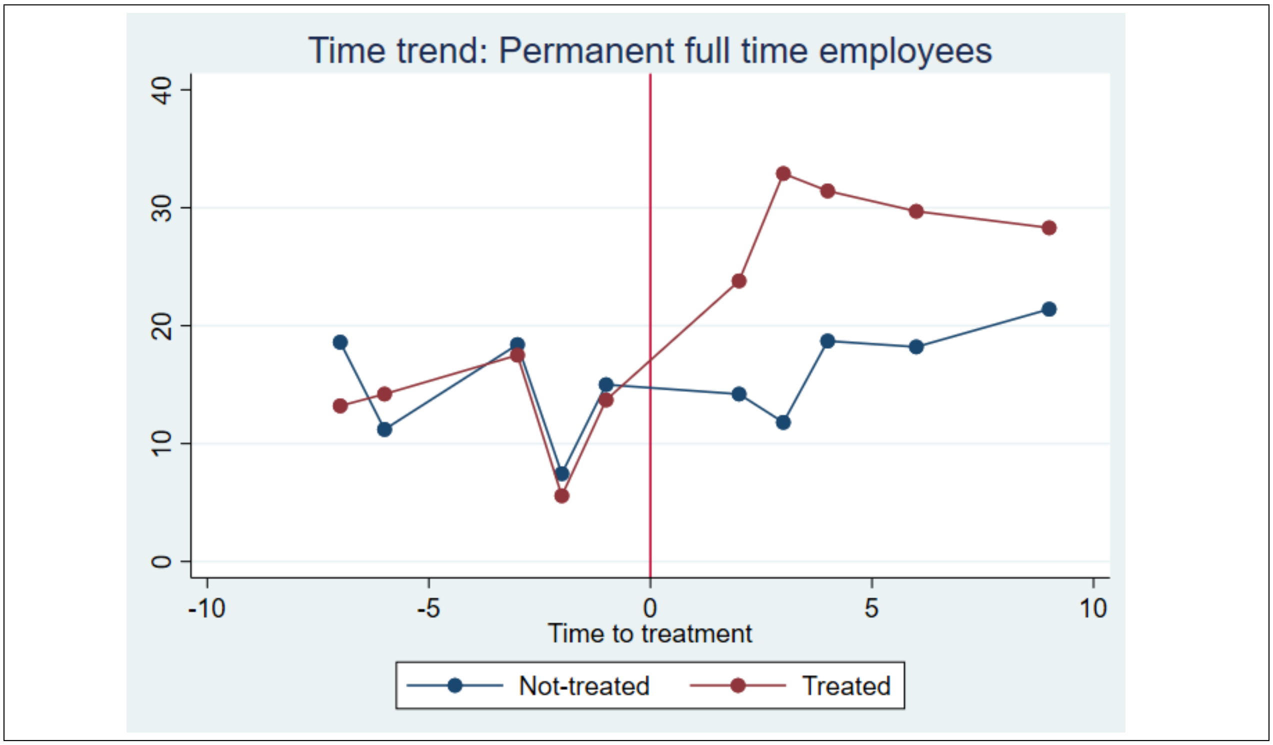


Figure 3-9: Time Trend in the Number of Permanent Full Time Employees (WBES).

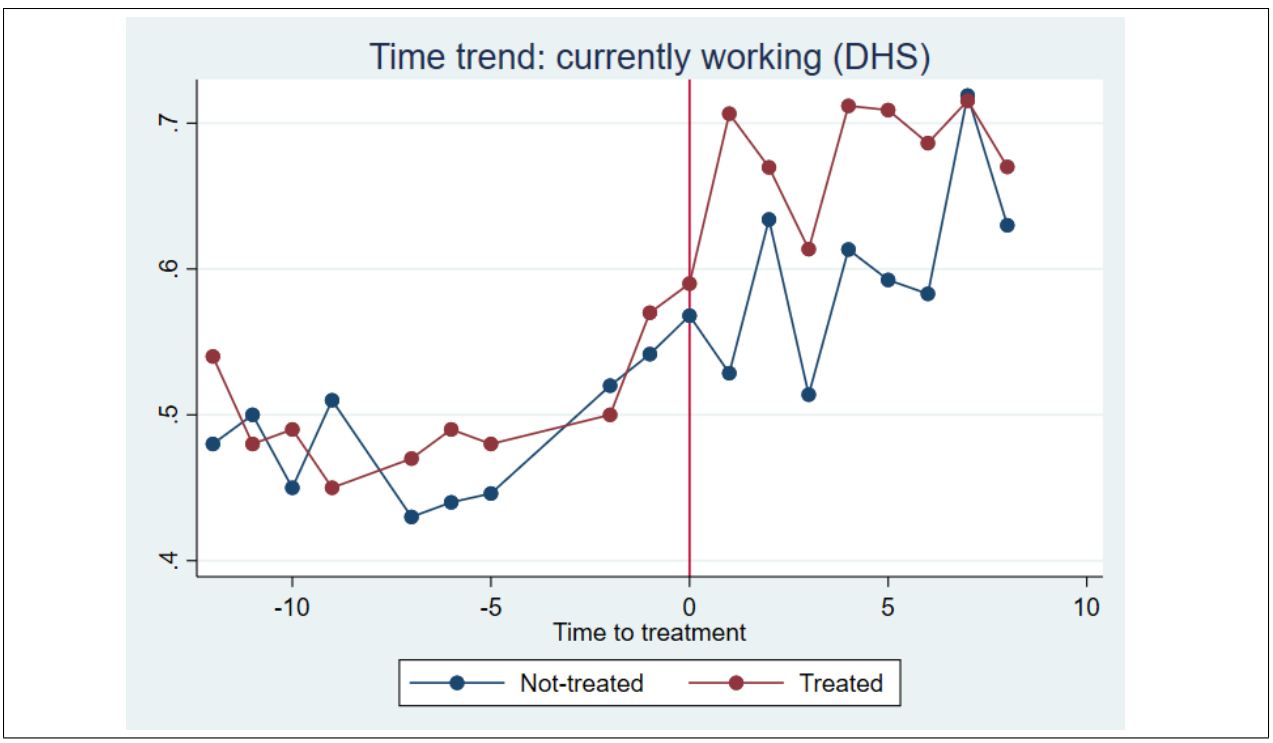


Figure 3-10: Time Trend in the Likelihood of being Employed at the Time of Survey (DHS).

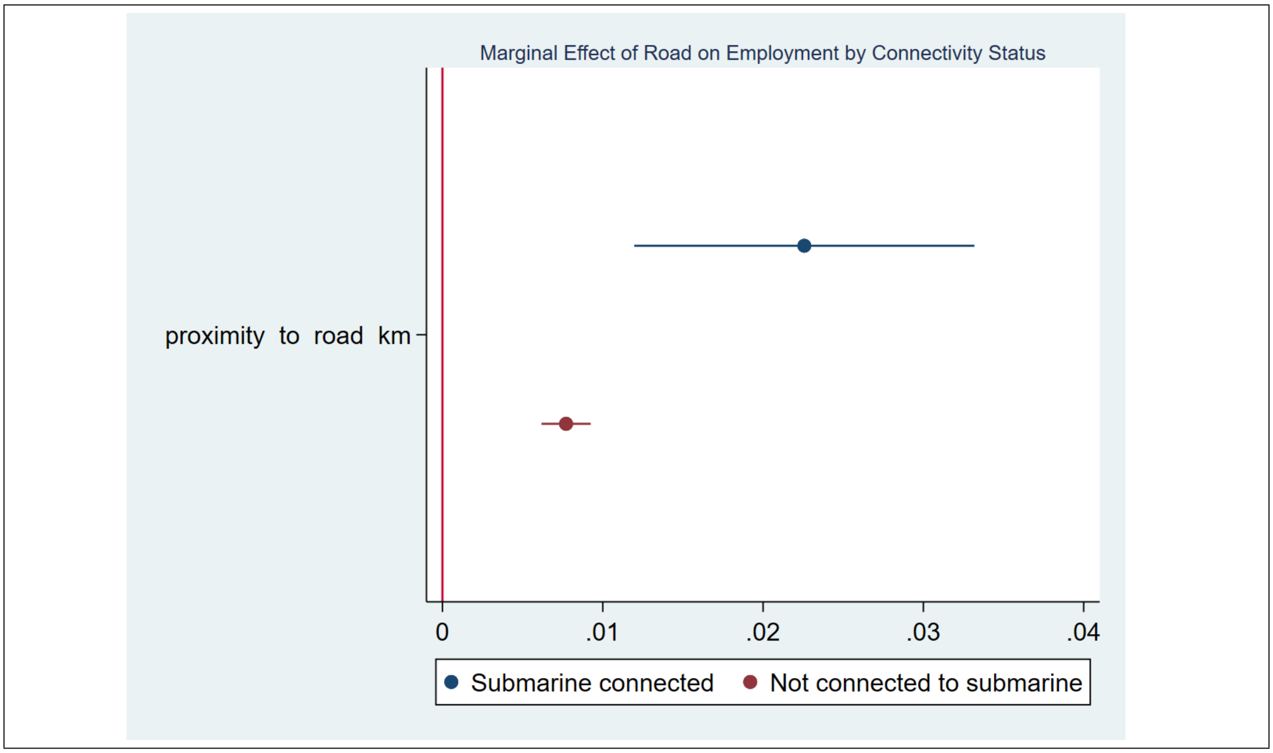


Figure 3-11: Road and Internet Effects on Likelihood to be Currently Working.

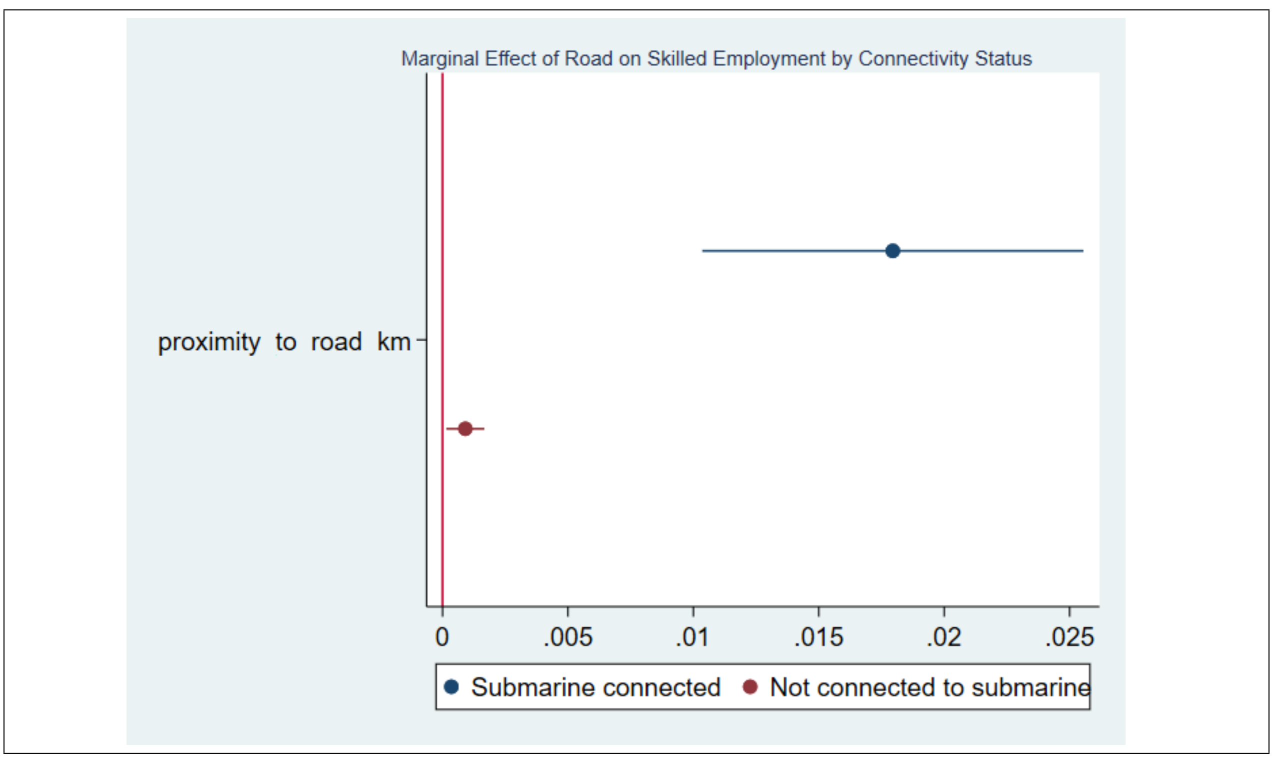


Figure 3-12: Road and Internet Effects on Skilled Employment.

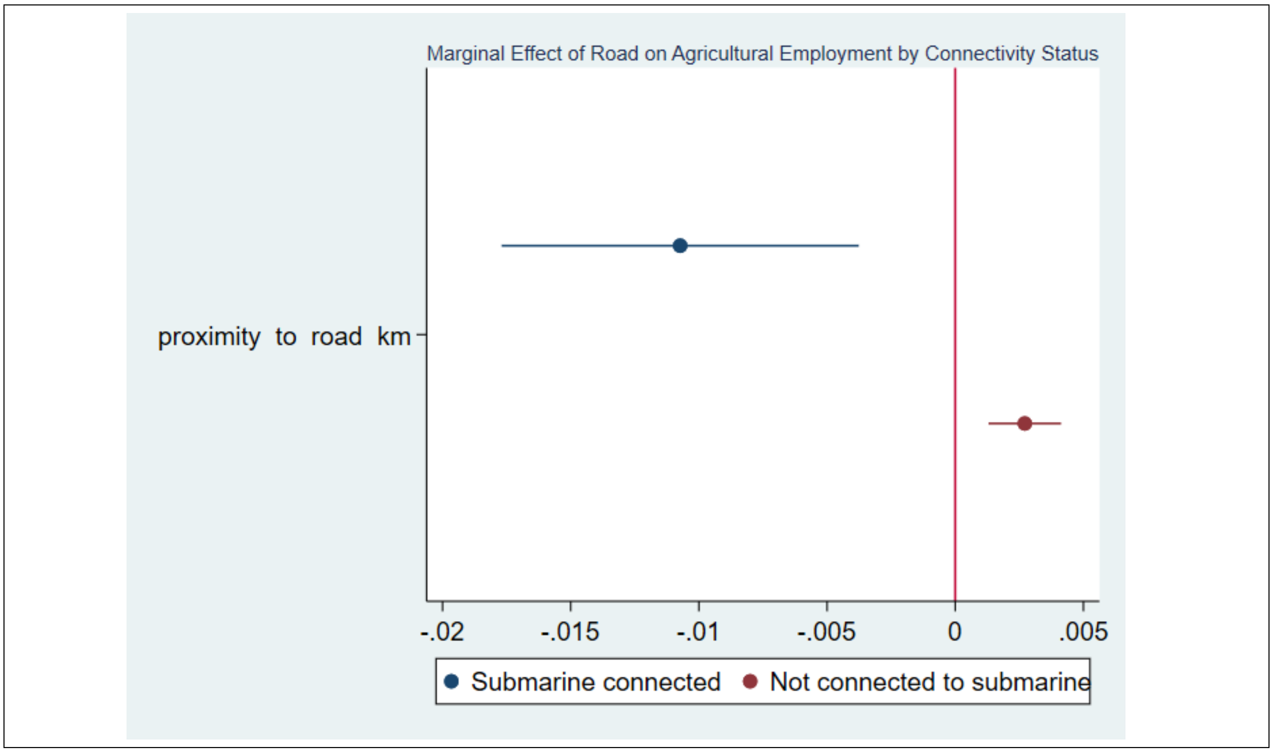


Figure 3-13: Road and Internet Effects on Agricultural Employment.

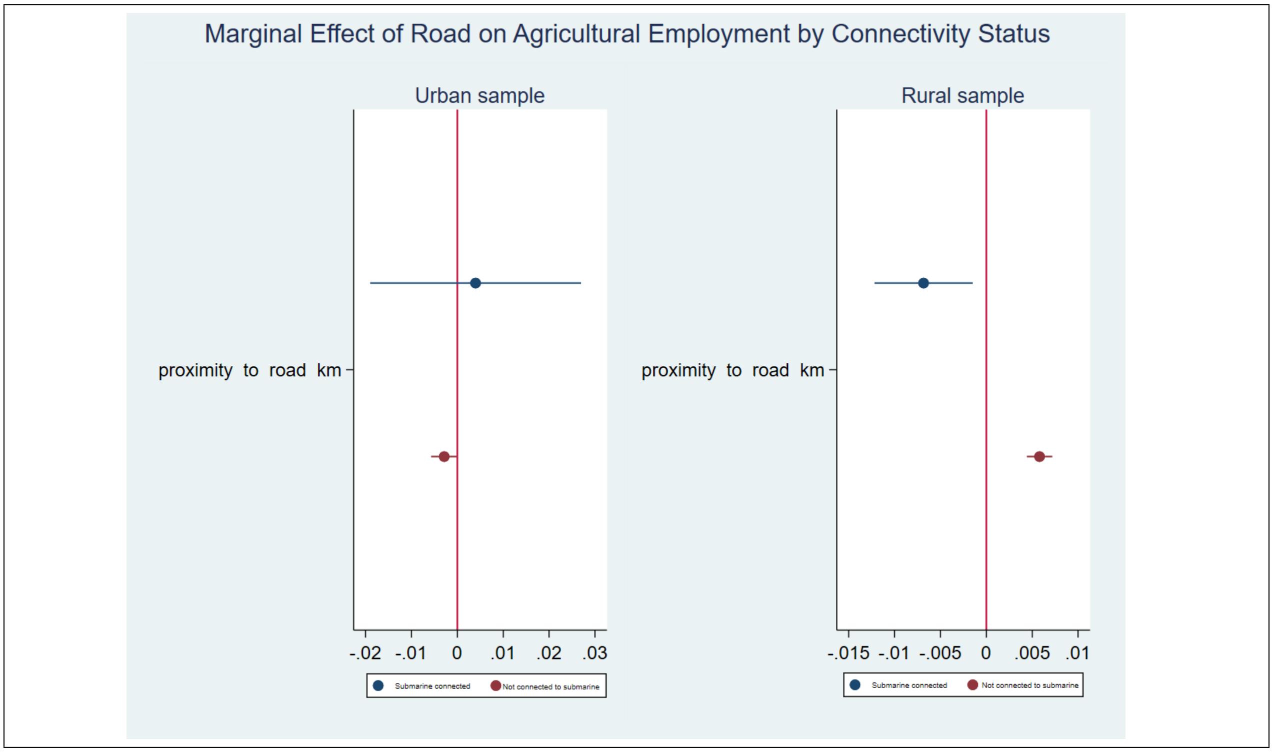


Figure 3-14: Road and Internet Effects on Agricultural Employment in Rural Areas.

Table 3-1: 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 3-2: Descriptive statistics: DHS Sample

	DHS				
Variables	Observation	Mean	Std.Dev	Min	Max
Submarine has arrived	780,422	0.4159	0.4929	0	1
Proximity to road (KM)	780,422	-9.702	17.289	-221.4656	-.0000708
Submarine (2 km) * proximity to road (KM)	780,422	-0.8802	4.3964	-192.9587	0
Proximity to Hypothetical road (KM)	780,422	-34.4518	34.2275	-273.9815	-.0010025
Submarine (2 km) * proximity to hypothetical road (KM)	780,422	-6.1458	18.4237	-221.7079	0
lightening	780,422	0.4987	0.3287	0	2.120674
Terrain elevation	780,422	679.2284	668.5519	-92	3979
slope	780,422	1.2647	1.7262	0	15.9138
Internet connected (0.5 m)	780,422	0.0084	0.0914	0	1
Internet connected (1 km)	780,422	0.0245	0.1545	0	1
Internet connected (1.5 km)	780,422	0.0451	0.2076	0	1
Internet connected (2 km)	780,422	0.4184	0.4933	0	1
Submarine Internet (0.5 km)	780,422	0.0027	0.0517	0	1
Submarine Internet (1 km)	780,422	0.0144	0.1191	0	1
Submarine Internet (1.5 km)	780,422	0.0201	0.1402	0	1
Submarine Internet (2 km)	780,422	0.1937	0.3952	0	1
Currently working	696,417	0.677	0.4676	0	1
Worked in the past year	756,335	0.7387	0.4393	0	1
Low skilled worker	724,778	0.4864	0.4984	0	1
Services	717,891	0.2399	0.4202	0	1
Skilled worker	752,131	0.196	0.392	0	1
High skilled worker	767,327	0.1091	0.3117	0	1
Self-employed farmer	752,131	0.2208	0.4148	0	1
Farmer (employee and self-emplyed)	752,131	0.3	0.4582	0	1
age	756,335	35.7444	7.9791	20	65
Female individual	756,335	0.6762	0.4679	0	1
Rural residence	756,335	0.6299	0.4828	0	1
No Education	756,335	0.4145	0.4926	0	1
Primary education	756,335	0.2686	0.4432	0	1
No education up to primary education	756,335	0.6831	0.4653	0	1
Secondary education and /or more	756,335	0.3169	0.4653	0	1
20-34 age group	627,697	0.4864	0.4998	0	1
35 to 44 age group	627,697	0.3405	0.4739	0	1
45 to 64 age group	627,697	0.1732	0.3784	0	1
Survey year	780,422			1998	2018

Table 3-3: Descriptive statistics: WBES Sample

WBES					
Variables	Observation	Mean	Std.Dev	Min	Max
Submarine has arrived	15,057	0.9443	0.2293	0	1
Proximity to road (KM)	15,028	-2.3084	4.6026	-55.85029	-.0005369
Submarine (2 km) * proximity to road (KM)	15,028	-0.296	1.2531	-15.97813	0
Proximity to Hypothetical road (KM)	15,028	-30.8241	30.1415	-146.8549	-.3167718
Submarine (2 km) * proximity to hypothetical road (KM)	15,028	-11.3393	23.0097	-104.9434	0
lightening	15,028	0.4507	0.3021	.0098068	1.687402
Terrain elevation	15,028	417.6429	556.265	1	2098
slope	15,028	0.5981	0.7434	0	4.41208
Internet connected (0.5 m)	15,028	0.1459	0.353	0	1
Internet connected (1 km)	15,057	0.2573	0.4372	0	1
Internet connected (1.5 km)	15,028	0.3189	0.4661	0	1
Internet connected (2 km)	15,028	0.7837	0.4117	0	1
Submarine Internet (0.5 km)	4,433	0.1041	0.3055	0	1
Submarine Internet (1 km)	15,057	0.1778	0.4372	0	1
Submarine Internet (1.5 km)	15,028	0.2573	0.3824	0	1
Submarine Internet (2 km)	15,057	0.3645	0.4813	0	1
Etablisment's age	14,531	15.004	12.77	0	168
Telecommunication is an obstable to operation	13,929	0.2685	0.4432	0	1
Losses as percentage of annual sales	12,139	9.2275	13.4941	0	160
Percentage owned by the governement	15,032	0.8456	6.9682	0	100
Secured a contract with government in last fiscal year	7,181	0.2041	0.4031	0	1
Hours of operation per week last year	15,057	35.4817	26.1427	10	84
Number of permanent full time employee at the end of 3 fiscal years ago	14,411	32.4492	178.1041	0	60
Number of full time skilled employee at the end of last fiscal years ago	13,653	15.731	26.5068	0	60
Number of full time male employee at the end of last fiscal years ago	6,730	16.171	29.117	0	60
Number of full time female employee at the end of last fiscal years ago	6,730	5.8948	20.8299	0	43
Number of full time employee last year	8,080	16.4357	40.7089	0	60
Number of full time unskilled employee at the end of last fiscal years ago	6,613	10.5485	36.5239	0	60
Survey year	15,033			2003	2018

Table 3-4: Descriptive statistics: LSMS Sample

		LSMS			
Variables	N	Mean	Std.Dev	Min	Max
Submarine has arrived	179,675	0.8925	0.3098	0	1
Proximity to road (KM)	179,675	-10.2243	15.2169	-192.6164	-0.0036
Submarine (2 km) * proximity to road (KM)	179,675	-0.0444	0.2836	-6.2234	0
Proximity to Hypothetical road (KM)	179,675	-26.9828	22.1511	-164.0193	-0.0205
Submarine (2 km) * proximity to hypothetical road (KM)	179,675	-0.8368	5.5502	-129.5013	0
lightening	179,675	0.3293	0.1204	0.0004	1.326
Terrain elevation	179,675	0.7763	0.4162	0.001	2.574
slope	179,675	6.7316	10.0516	0	110.6
Internet connected (0.5 m)	179,675	0.0004	0.1976	0	1
Internet connected (1 km)	179,675	0.0039	0.1935	0	1
Internet connected (1.5 km)	179,675	0.0275	0.1636	0	1
Internet connected (2 km)	179,675	0.0424	0.2015	0	1
Submarine Internet (0.5 km)	179,675	0.0021	0.0454	0	1
Submarine Internet (1 km)	179,675	0.0021	0.0454	0	1
Submarine Internet (1.5 km)	179,675	0.0275	0.1636	0	1
Submarine Internet (2 km)	179,675	0.0415	0.1994	0	1
age	179,675	38.517	12.5809	20	65
Female individual	179,675	0.5324	0.499	0	1
Rural residence	179,675	0.8329	0.3731	0	1
No Education	179,675	0.5248	0.4994	0	1
Primary Education	179,675	0.1628	0.3691	0	1
Secondary Education	179,675	0.6874	0.4636	0	1
Higher Education	179,675	0.2095	0.407	0	1
Hours worked in the Last seven day	179,675	12.6073	75.8975	0	105
Survey years	179,675			2005	2019

Table 3-5: Balance Check: Differences in Employment Outcomes Before Submarine Internet Arrival Between Connected and non-Connected Observations

	Connected **Std.Dev**	**Not connected** **Std.Dev**	**Difference** **t-stat**
DHS			
Currently working	0.57 0.5	0.57 0.5	0 1.54
Worked during last year	0.71 0.45	0.72 0.45	0.01 1.41
Low skilled worker	0.63 0.48	0.56 0.50	-0.07 -3.95
Skilled worker	0.24 0.42	0.15 0.34	-0.09 -9.31
High skilled worker	0.14 0.27	0.08 0.35	-0.06 -8.25
Farming is main employment	0.15 0.36	0.29 0.45	0.15 1.27
Services	0.27 0.44	0.18 0.39	-0.09 -8.54
WBES			
Number of full time employee at business start	10.65 12.65	12.04 13.81	1.38 1.79
Number of full time employees last fiscal year	14.91 31.58	17.29 37.45	2.38 0.61
Number of female full time employee last fiscal year	5.39 23.03	6.26 22.03	0.87 0.45
Number of skilled full time employee last fiscal year	13.84 22.66	19.93 29.24	6.09 0.64
Telecommunication is an obstacle to operations	0.29 0.44	0.27 0.46	-0.02 -2.71
Losses as percentage of annual sales	8.95 14.15	8.58 14.68	-0.37 -1.38
Percentage own by the government	0.73 7.32	0.75 6.98	0.02 0.12
Secured a contract with government in last fiscal year	0.20 0.40	0.14 0.35	-0.06 -5.57
Hours of operations per week in last fiscal year	25.67 29.35	32.23 34.49	6.56 12.01
LSMS			
Hours worked in last 7 days	6.15 17.06	10.03 23.64	3.88 4.05

Table 3-6: First Stage

Variables	(1) DHS Proximity to road (Km)	(2) DHS Submarine internet * road	(3) WBES Proximity to road (Km)	(4) WBES Submarine internet * road	(5) LSMS Proximity to road (Km)	(6) LSMS Submarine internet * road
Proxymity to hyp. roads	0.0369***	0.0123	0.154***	-5.86e-13	0.0022***	-0.0012***
	(0.00929)	(0.0546)	(0.00371)	(4.20e-13)	(0.0003)	(0.000002)
Submarine * hyp. roads	0.000693*	0.0580***	0.0319	3.815***	-0.0548***	0.0284***
	(0.000409)	(0.00524)	(0.0243)	(3.23e-08)	(0.0009)	(0.0003)
Lightning	0.109	-0.00853***	0.731***	-0.0691***	-2.3967***	-0.1188***
	(0.398)	(0.000921)	(0.174)	(7.70e-10)	(0.1754)	(0.0079)
Terrain elevation	-0.0149*	0.0000475	-0.00242***	-0.0428***	-0.0906***	-0.0182***
	(0.00860)	(0.0000686)	(0.000397)	(9.41e-11)	(0.0248)	(0.0019)
N	780422	780422	15028	15028	179675	179675
ymean	-9.445	-0.882	-0.441	-0.214	-8.323	-0.0618
r2	0.275	0.171	0.992	1	0.858	0.720
r2_a	0.275	0.171	0.992	1	0.858	0.720
Kleibergen-Paap F-statistic	46.03	46.03	33.83	33.83	60.15	60.15
$P > F$	0	0	0.2747	0.2747	2.70e-12	2.70e-12
Cragg-Donald Wald F-statistic	792.3	792.3	218.478	218.478	57.57	57.57
Windmeijer F-statistic	33.37	33.37	30.43	30.43	36.29	36.29
Kleibergen-Paap Wald F-statistic	7.875	7.875	33.831	33.831	18.94	18.94
Anderson-Rubin Wald F-statistic	6.673	6.673	41.81	41.81	37.41	37.41
Age	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-7: The Complementary Effects of Internet and Roads on Employment

	(1)	(2)	(3)	(4)	(5)	(6)
	DHS		**LSMS**		**WBES**	
Variables	**Worked during last year**	**currently working**	**Hours Worked last week**	**Hours of operation per week in last year**	**Number of permanent full time employees 3 fiscal years ago**	**Number of full time employees last year**
Submarine internet	0.0833***	0.0568**	-0.341***	0.575**	0.560***	0.973***
	(0.0236)	(0.0219)	(0.038)	(0.240)	(0.0721)	(0.193)
Proximity to road (Km)	0.00779***	0.00787***	0.959	3.782***	1.445***	0.132***
	(0.000478)	(0.000470)	(0.646)	(0.0602)	(0.433)	(0.0449)
Submarine internet * road	0.0167**	0.0116*	3.491***	0.0934***	0.0268***	0.0727***
	(0.00522)	(0.00475)	(0.839)	(0.0246)	(0.00298)	(0.0181)
N	756335	696417	179675	14526	11294	6633
ymean	0.744	0.680	12.26	35.28	12.00	15.235
R-sq	0.0343	0.0654	-0.1233	0.331	0.158	0.0413
R-sq_a	0.0342	0.0653	-0.0932	0.327	0.153	0.0309
Kleibergen-Paap F-statistic	79.65	81.52	17.72	1.094	1.178	1.463
$P > F$	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	848.5	791.4	139.8	1282.5	925.1	411.6
Windmeijer F-statistic	28.8	25.32	48.37	25.39	18.95	51.5
Kleibergen-Paap Wald F-statistic	18.72	19.21	3.427	25.39	51.50	3984.6
Age	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✗	✗	✗
Place of residence	✓	✓	✓	✗	✗	✗
Education	✓	✓	✓	✗	✗	✗
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-8: The Complementary Effects of Internet and Roads on Employment by Sector of Occupation

	(1)	(2)	(3)	(4)	(5)	(6)
	DHS				WBES	
	Low skilled worker	**Services**	**Skilled worker**	**High skilled worker**	**Number of full time skilled employees at end of last fiscal year**	**Number of full time unskilled employee at end of last fiscal year**
Submarine internet	-0.131***	0.168***	0.129***	0.0962***	0.428***	0.564***
	(0.0195)	(0.0236)	(0.0189)	(0.0143)	(0.0438)	(0.0822)
Proximity to road (KM)	0.00524***	0.00126***	0.000939***	0.000988	0.976***	0.563***
	(0.000411)	(0.000283)	(0.000233)	(0.00176)	(0.127)	(0.209)
Submarine internet * road	-0.0194***	0.0294***	0.0218***	0.0156***	0.0372***	0.0621***
	(0.00422)	(0.00529)	(0.00417)	(0.00314)	(0.00693)	(0.0158)
N	724778	717891	752131	767327	12276	3306
ymean	0.528	0.226	0.187	0.121	15.942	8.584
R-sq	0.139	0.0956	0.0734	0.0471	0.115	0.0914
R-sq_a	0.139	0.0955	0.0733	0.0471	0.109	0.0718
Kleibergen-Paap F-statistic	81.24	81.55	80.89	81.15	1.156	1.280
$P > F$	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	811.4	808.3	885.1	850.6	1039.6	180.3
Windmeijer F-statistic	17.52	20.01	26.87	16.71	51.5	41.33
Kleibergen-Paap Wald F-statistic	19.18	19.23	18.99	19.14	41.33	66.88
Age	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✗	✗
Place of residence	✓	✓	✓	✓	✗	✗
Education	✓	✓	✓	✓	✗	✗
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$*. We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-9: The Complementary Effects of Internet and Roads on Agricultural Employment

	(1)	(2)	(3)	(4)
	All sample		**Rural sample**	
	Farming self-employed	**Farming employee and self-employed**	**Farming self-employed**	**Farming employee and self-employed**
Submarine internet	-0.147***	-0.112***	-0.104***	-0.128***
	(0.0395)	(0.0363)	(0.0327)	(0.0353)
Proximity to road (KM)	0.00209**	0.00270***	0.00312***	0.00601***
	(0.000842)	(0.000802)	(0.000610)	(0.000710)
Submarine internet * road	-0.0284***	-0.0263***	-0.0133***	-0.0169***
	(0.00804)	(0.00779)	(0.00397)	(0.00424)
N	752131	752131	465918	465918
ymean	0.208	0.284	0.304	0.415
r2	0.253	0.220	0.332	0.165
r2_a	0.253	0.220	0.331	0.164
Kleibergen-Paap F-statistic	45.74	45.74	64.41	64.41
$P > F$	0	0	0	0
Cragg-Donald Wald F-statistic	885.1	885.1	1219.3	1219.3
Windmeijer F-statistic	35.21	41.74	30.01	31.85
Kleibergen-Paap Wald F-statistic	7.590	7.590	12.39	12.39
Age	✓	✓	✓	✓
Gender	✓	✓	✓	✓
Place of residence	✓	✓	✗	✗
Education	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-10: Heterogeneous Effects of the Complementary of Internet and Roads on Employment: Place of Residence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Worked during past year				**Currently working**						**Hours worked**
	Urban Sample	**Rural Sample**	**high urbanized Sample**	**Low urbanized Sample**	**Urban Sample**	**Rural Sample**	**high urbanized Sample**	**Low urbanized Sample**	**Urban Sample**	**Rural Sample**	**high urbanized Sample**	**Low urbanized Sample**
				DHS							**LSMS**	
Submarine internet	0.210*	0.0369	0.0891	0.0603**	0.190*	0.0165	0.0419	0.0514*	-2.753***	0.0243	-0.340***	-2.826***
	(0.101)	(0.0239)	(0.0516)	(0.0209)	(0.0920)	(0.0240)	(0.0443)	(0.0209)	(0.239)	(0.185)	(0.0381)	(0.655)
Proximity to road (KM)	0.00622***	0.00822***	0.00962***	0.00689***	0.00709***	0.00758***	0.00938***	0.00725***	3.805***	28.38	0.960	2.175
	(0.00118)	(0.000613)	(0.00101)	(0.000515)	(0.00119)	(0.000581)	(0.00104)	(0.000509)	(0.947)	(22.39)	(0.646)	(1.534)
Submarine internet * road	0.0905*	0.00228	0.0152	0.0122**	0.0797*	-0.000414	0.00702	0.00926*	7.098***	45.38	3.491***	12.03***
	(0.0442)	(0.00312)	(0.0123)	(0.00425)	(0.0400)	(0.00306)	(0.0104)	(0.00416)	(0.878)	(37.99)	(0.839)	(1.614)
N	287502	468833	295535	460800	265370	431047	263715	432702	23599	156076	179675	84925
ymean	0.734	0.752	0.764	0.733	0.682	0.696	0.671	0.671	14.60	10.75	12.61	12.57
r2	-0.318	0.0754	0.0151	0.0675	-0.239	0.117	0.0767	0.0816	0.00632	-0.365	-0.123	-10.44
r2_a	-0.318	0.0753	0.0150	0.0675	-0.240	0.117	0.0767	0.0816	0.00687	-0.365	-0.123	-10.44
F-Statistic	421.9	512.0	986.1	645.3	418.5	649.9	982.8	752.1	57.05	46.62	170.7	36.62
$P > F$	0	0	0	0	0	0	0	0	0	0	0.0000166	1.08e-180
Kleibergen-Paap F-statistic	25.81	67.77	21.60	73.31	24.89	69.03	22.38	74.27	10.36	3.81	64.03	2.55
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donal Wald F-statistic	47.89	1146.8	125.3	1050.5	45.69	1048.4	114.9	1002.2	330.9	152.2	327.0	331.5
Kleibergen-Paap Wald F-statistic	5.120	15.50	4.382	17.27	4.940	15.79	4.532	17.54	25.40	7.466	126.5	5.090
Windmeijer F-statistic	7.436	17.88	57.84	20.62	7.586	9.036	26.38	29.77	4.603	0.755	1.261	1.257
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✗	✗	✓	✓	✗	✗	✓	✓	✗	✗	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-11: Heterogeneous Effects of the Complementary of Internet and Roads on Employment: Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DHS				LSMS		WBES	
	Worked during past year		Currently working		Hours worked		Number of full time male employee at end of last fiscal year	
	Male Sample	Female Sample	Male Sample	Female Sample	Male Sample	Female Sample	Male Sample	Female Sample
Submarine internet	0.0589**	0.0964***	0.0393	0.0830**	-1.127**	-0.449*	1.078***	0.131***
	(0.0209)	(0.0287)	(0.0205)	(0.0275)	(0.4011)	(0.0360)	(0.327)	(0.0478)
Proximity to road (KM)	0.00410***	0.00899***	0.00462***	0.00879***	0.356***	1.011***	0.448***	1.004***
	(0.000467)	(0.000574)	(0.000597)	(0.000551)	(0.0392)	(0.152)	(0.0392)	(0.165)
Submarine internet * road	0.0118*	0.0193**	0.00527	0.0163**	0.833***	0.041***	0.0916***	0.00157
	(0.00489)	(0.00610)	(0.00459)	(0.00582)	(0.179)	(0.071)	(0.0337)	(0.00273)
N	245289	511046	176813	519604	94750	130794	5882	5043
ymean	0.898	0.673	0.856	0.620	12.64	10.42	13.332	5.326
r2	0.0161	-0.0155	0.0317	0.0234	-0.0117	-0.417	0.0426	0.0278
r2_a	0.0159	-0.0156	0.0315	0.0233	-0.0115	-0.417	0.0311	0.0141
Kleibergen-Paap F-statistic	58.36	83.40	60.52	85.79	1.485	0.963	1.485	0.963
$P > F$	0	0	0	0	0	0	0	0
Cragg-Donal Wald F-statistic	290.0	558.1	230.5	560.2	120.3	163.0	391.7	321.0
Kleibergen-Paap Wald F-statistic	13.09	19.77	13.41	20.44	182.9	168.7	3000.9	5.997
Windmeijer F-statistic	23.42	42.48	28.37	30.71	142.0	458.6	142.0	458.6
Age	✓	✓	✓	✓	✓	✓	✓	✓
Gender	✗	✗	✗	✗	✗	✗	✗	✗
Place of residence	✓	✓	✓	✓	✓	✓	✗	✗
Education	✓	✓	✓	✓	✓	✓	✗	✗
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-12: Heterogeneous Effects of the Complementary of Internet and Roads on Employment: Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Worked during past year				Currently working				Hours worked			
	Not educated Sample	Low educated Sample (Primary)	Not more than primary Sample	higher educated Sample	Not educated Sample	Low educated Sample (Primary)	Not more than primary Sample	higher educated Sample	Not educated Sample	Low educated Sample (Primary)	Not more than primary Sample	higher educated Sample
	DHS								LSMS			
Submarine internet	0.181**	0.0677**	0.106***	0.0629*	0.146**	0.0510*	0.0846**	0.0566*	0.542***	-0.395***	-0.483***	-0.607***
	(0.0599)	(0.0248)	(0.0302)	(0.0279)	(0.0542)	(0.0256)	(0.0288)	(0.0272)	(0.0617)	(0.0310)	(0.124)	(0.0772)
Proximity to road (KM)	0.00649***	0.00939***	0.00770***	0.00701***	0.00619***	0.00985***	0.00749***	0.00784***	5.833	4.101***	3.708***	1.366***
	(0.000475)	(0.000963)	(0.000473)	(0.00105)	(0.000468)	(0.00101)	(0.000460)	(0.00115)	(4.399)	(0.679)	(1.241)	(0.2934)
Submarine internet * road	0.0281**	0.0101*	0.0157**	0.0206*	0.0223*	0.00849	0.0125*	0.0168*	2.870	6.304**	1.863	6.982**
	(0.00992)	(0.00468)	(0.00537)	(0.00846)	(0.00889)	(0.00474)	(0.00506)	(0.00808)	(2.808)	(2.583)	(1.646)	(3.332)
N	295558	196514	492072	264263	278239	180856	459095	237322	18800	149588	27240	58901
ymean	0.740	0.769	0.752	0.732	0.671	0.703	0.683	0.675	14.19	10.89	18.98	15.25
r2	0.0421	0.0362	0.0461	0.0811	0.0914	0.0572	0.0832	0.0914	-0.237	-0.242	-1.427	-0.367
r2_a	0.0419	0.0359	0.0460	0.0809	0.0912	0.0570	0.0831	0.0912	-0.236	-0.242	-1.426	-0.366
Kleibergen-Paap F-statistic	73.58	51.84	75.90	56.00	73.87	53.23	77.11	53.32	3.161	10.71	9.055	9.315
P ¿ F	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donal Wald F-statistic	213.5	345.9	595.0	228.5	204.3	323.8	561.9	205.6	48.37	180.0	73.95	53.13
Kleibergen-Paap Wald F-statistic	17.43	11.21	17.94	11.90	17.56	11.58	18.28	11.25	40.91	186.5	43.16	51.41
Windmeijer F-statistic	13.82	23.62	26.33	24.79	14.43	21.67	16.04	28.69	22.71	250.5	50.90	102.5
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-13: Heterogeneous Effects of the Complementary of Internet and Roads on Employment: Age

	(1)	(2)	(4)	(5)	(7)	(8)	(9)	(11)	(12)
		Worked during past year			Currently working			Hours worked	
	20-34 groupe age Sample	35 -44 groupe age Sample	45-64 groupe age Sample	20-34 groupe age Sample	35 -44 groupe age Sample	45-64 groupe age Sample	20-34 groupe age Sample	35 -44 groupe age Sample	45-64 groupe age Sample
	DHS						LSMS		
Submarine internet	0.0923**	0.0608*	0.0808**	0.0560*	0.0492*	0.0584*	-0.607***	-0.0316	-0.0294
	(0.0283)	(0.0240)	(0.0288)	(0.0252)	(0.0242)	(0.0291)	(0.0772)	(0.0221)	(0.464)
Proximity to road (KM)	0.00799***	0.00796***	0.00638***	0.00794***	0.00810***	0.00706***	1.366***	0.7445***	0.277***
	(0.000520)	(0.000521)	(0.000623)	(0.000496)	(0.000537)	(0.000695)	(0.2934)	(0.103)	(0.0488)
Submarine internet * road	0.0201**	0.0107*	0.0141*	0.0128*	0.00880	0.0107	1.849***	1.025***	1.017
	(0.00646)	(0.00491)	(0.00606)	(0.00565)	(0.00489)	(0.00598)	(0.350)	(0.139)	(0.682)
N	451922	201169	103244	420560	186046	89811	58901	56848	60527
ymean	0.692	0.813	0.844	0.626	0.753	0.785	15.25	11	11.83
r2	0.0253	0.0308	0.0484	0.0601	0.0626	0.0686	-0.367	-0.0110	-0.0553
r2_a	0.0251	0.0306	0.0480	0.0600	0.0623	0.0681	-0.366	-0.0107	-0.0550
Kleibergen-Paap F-statistic	77.79	76.70	50.13	79.84	78.45	50.85	7.427	7.935	6.247
$P > F$	0	0	0	0	0	0	0	0	0
Cragg-Donal Wald F-statistic	445.4	283.5	119.0	419.2	264.0	107.6	346.5	459.5	434.2
Kleibergen-Paap Wald F-statistic	18.19	18.10	11.03	18.70	18.55	11.22	217.8	633.6	209.4
Windmeijer F-statistic	70.19	13.35	11.20	64.69	11.04	8.539	53.13	152.4	49.91
Age	✗	✗	✗	✗	✗	✗	✗	✗	✗
Gender	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-14: Effects of the Complementary of Internet and Roads on Firms Operations and Productivity

	(1)	(2)	(3)	(4)
			WBES	
Variables	Telecommunication is an obstacle to operation	Losses as percentage of annual sales	percentage owned by the governemnt	Secured a contract with governemnt in the last fiscal year
Submarine internet	-0.0678***	-2.293***	1.268***	0.172***
	(0.00976)	(0.267)	(0.0477)	(0.0152)
Proximity to road (KM)	0.0362***	-0.712***	-0.295***	-0.00146
	(0.00115)	(0.0248)	(0.00792)	(0.00421)
Submarine internet * road	-0.00361***	-0.146***	0.0847***	0.0111***
	(0.00129)	(0.0451)	(0.00254)	(0.00365)
N	13473	11742	14525	6768
ymean	0.270	9.322	0.757	0.183
r2	0.139	0.0949	0.0301	0.167
r2_a	0.135	0.0892	0.0252	0.160
Kleibergen-Paap F-statistic	1.096	1.059	1.094	1.362
$P > F$	0	0	0	0
Cragg-Donald Wald F-statistic	1181.8	1051.7	1282.4	394.2
Windmeijer F-statistic	25.59	18.95	25.39	12.39
Kleibergen-Paap Wald F-statistic	25.99	18.95	25.39	183.9
Age of the firm	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 * *p < .05 * * * p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-15: Heterogeneous Effects of the Complementary of Internet and Roads on Skilled Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							High skilled worker							
	All Sample	**Urban Sample**	**Rural Sample**	**Male Sample**	**Female Sample**	**Not educated Sample**	**Low educated Sample (Primary)**	**Not more than primary Sample**	**higher educated Sample**	**20-34 groupe age Sample**	**35 -44 groupe age Sample**	**45-64 groupe age Sample**	**More developed Sample**	**Less developed Sample**
Submarine Internet	0.0962***	0.0807*	0.0612***	0.0669***	0.114***	0.152***	0.0517***	0.0970***	0.0516**	0.111***	0.0817***	0.0611**	0.200**	0.0620***
	(0.0143)	(0.0359)	(0.0126)	(0.0176)	(0.0165)	(0.0324)	(0.0112)	(0.0149)	(0.0189)	(0.0175)	(0.0166)	(0.0230)	(0.0612)	(0.0111)
Proximity to road (KM)	0.0000988	0.00128**	-0.000410*	0.000206	0.0000295	0.0000358	-0.000593*	-0.0000632	-0.0000256	0.000250	-0.0000404	-0.0000714	-0.00107	0.000377*
	(0.000176)	(0.000474)	(0.000164)	(0.000318)	(0.000180)	(0.000137)	(0.000289)	(0.000117)	(0.000734)	(0.000197)	(0.000257)	(0.000362)	(0.000556)	(0.000191)
Submarine Internet * road	0.0156***	0.0203	0.00609***	0.0115**	0.0183***	0.0218***	0.00735***	0.0137***	0.00924	0.0189***	0.0112***	0.0118*	0.0426**	0.00784***
	(0.00314)	(0.0153)	(0.00160)	(0.00404)	(0.00349)	(0.00535)	(0.00198)	(0.00264)	(0.00543)	(0.00402)	(0.00320)	(0.00472)	(0.0148)	(0.00216)
N	767327	291558	475769	245304	522023	300363	199363	499726	267601	459291	203919	104117	300383	466944
ymean	0.121	0.209	0.067	0.166	0.100	0.034	0.070	0.048	0.257	0.115	0.129	0.133	0.140	0.109
r2	0.0471	0.0352	0.0449	0.0801	0.0114	-0.140	0.0332	-0.0176	0.0550	0.0174	0.0878	0.0960	-0.220	0.0810
r2_a	0.0471	0.0350	0.0448	0.0799	0.0113	-0.140	0.0329	-0.0177	0.0548	0.0173	0.0876	0.0956	-0.220	0.0810
Kleibergen-Paap F-statistic	81.15	25.38	68.70	58.25	85.42	73.52	51.83	75.86	56.02	79.25	78.44	51.29	22.25	74.15
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	850.6	48.40	1142.5	290.5	559.9	213.6	346.5	595.8	229.0	447.1	282.4	120.4	124.6	1058.4
Windmeijer F-statistic	16.71	11.97	6.50	7.77	8.05	6.40	9.00	13.75	6.93	8.80	6.75	18.91	6.50	7.54
Kleibergen-Paap Wald F-statistic	19.14	5.036	15.73	13.06	20.34	17.41	11.21	17.93	11.91	18.57	18.57	11.31	4.513	17.52
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-16: Heterogeneous Effects of the Complementary of Internet and Roads on Skilled jobs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Skilled worker							
	All Sample	Urban Sample	Rural Sample	Male Sample	Female Sample	Not educated Sample	Low educated Sample (Primary)	Not more than primary Sample	higher educated Sample	20-34 groupe age Sample	35 -44 groupe age Sample	45-64 groupe age Sample	More developed Sample	Less developed Sample
Submarine internet	0.129***	0.138*	0.0991***	0.171***	0.109***	0.199***	0.0862***	0.146***	0.0607**	0.133***	0.123***	0.116***	0.230**	0.100***
	(0.0189)	(0.0571)	(0.0179)	(0.0341)	(0.0161)	(0.0419)	(0.0185)	(0.0226)	(0.0214)	(0.0218)	(0.0209)	(0.0301)	(0.0715)	(0.0155)
Proximity to road (KM)	0.000939***	0.00296***	-0.000293	0.00286***	0.000141	0.000338	0.00146**	0.000484*	0.00234**	0.00106***	0.000929**	0.000699	-0.0000496	0.00114***
	(0.000233)	(0.000676)	(0.000230)	(0.000477)	(0.000216)	(0.000197)	(0.000474)	(0.000192)	(0.000801)	(0.000261)	(0.000300)	(0.000456)	(0.000664)	(0.000251)
Submarine internet * road	0.0218***	0.0429	0.00985***	0.0352***	0.0165***	0.0280***	0.0139***	0.0211***	0.0128*	0.0234***	0.0185***	0.0224***	0.0495**	0.0131***
	(0.00417)	(0.0244)	(0.00230)	(0.00818)	(0.00336)	(0.00690)	(0.00347)	(0.00405)	(0.00619)	(0.00495)	(0.00418)	(0.00624)	(0.0172)	(0.00310)
N	752131	286213	465918	244272	507859	295085	195580	490665	261466	448254	201003	102874	292000	460131
ymean	0.187	0.297	0.121	0.295	0.136	0.085	0.155	0.113	0.329	0.184	0.194	0.194	0.216	0.170
r2	0.0734	0.0278	0.0589	0.0533	0.0430	-0.0758	0.0703	0.00778	0.0793	0.0561	0.115	0.108	-0.170	0.122
r2_a	0.0733	0.0276	0.0588	0.0531	0.0429	-0.0760	0.0700	0.00768	0.0791	0.0560	0.115	0.108	-0.170	0.122
Kleibergen-Paap F-statistic	80.89	23.87	70.07	58.19	85.11	74.85	52.47	76.84	54.43	78.63	77.85	51.02	21.37	74.14
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	885.1	45.01	1219.3	288.8	596.1	247.4	347.0	632.1	229.5	471.4	289.5	121.8	134.8	1043.2
Windmeijer F-statistic	26.87	8.43	16.15	14.95	19.81	21.51	7.26	18.00	3.20	15.77	11.16	5.67	7.95	13.65
Kleibergen-Paap Wald F-statistic	18.99	4.748	16.04	13.04	20.12	17.78	11.34	18.10	11.56	18.33	18.39	11.24	4.320	17.49
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 ** p < .05 *** p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-17: Heterogeneous Effects of the Complementary of Internet and Roads on Unkilled Employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Low skilled worker							
	All Sample	**Urban Sample**	**Rural Sample**	**Male Sample**	**Female Sample**	**Not educated Sample**	**Low educated Sample (Primary)**	**Not more than primary Sample**	**higher educated Sample**	**20-34 groupe age Sample**	**35 -44 groupe age Sample**	**45-64 groupe age Sample**	**More developed Sample**	**Less developed Sample**
Submarine internet	-0.131***	0.0779	-0.102***	-0.225***	-0.0888***	-0.211***	-0.0565**	-0.132***	0.00154	-0.129***	-0.133***	-0.134***	-0.243***	-0.0562**
	(0.0195)	(0.0629)	(0.0244)	(0.0301)	(0.0213)	(0.0531)	(0.0217)	(0.0251)	(0.0203)	(0.0212)	(0.0243)	(0.0329)	(0.0606)	(0.0182)
Proximity to road (KM)	0.00524***	0.00310***	0.00479***	0.000397	0.00695***	0.00409***	0.00645***	0.00541***	0.000561	0.00545***	0.00530***	0.00388***	0.00520***	0.00475***
	(0.000411)	(0.000839)	(0.000497)	(0.000603)	(0.000482)	(0.000426)	(0.000915)	(0.000426)	(0.000823)	(0.000437)	(0.000508)	(0.000604)	(0.000937)	(0.000443)
Submarine internet * road	-0.0194***	0.0455	-0.0114***	-0.0356***	-0.0121**	-0.0270**	-0.00622	-0.0162***	0.00241	-0.0196***	-0.0187***	-0.0192**	-0.0419**	-0.00476
	(0.00422)	(0.0270)	(0.00316)	(0.00719)	(0.00439)	(0.00887)	(0.00381)	(0.00440)	(0.00578)	(0.00471)	(0.00479)	(0.00680)	(0.0144)	(0.00348)
N	724778	273255	451523	203643	521135	289648	189254	478902	245876	435086	193759	95933	277048	447730
ymean	0.529	0.640	0.462	0.659	0.473	0.408	0.495	0.443	0.692	0.567	0.469	0.476	0.583	0.495
r2	0.139	-0.0236	0.232	0.256	0.125	0.216	0.133	0.180	0.0676	0.123	0.143	0.202	0.0260	0.167
r2_a	0.139	-0.0238	0.232	0.256	0.125	0.216	0.133	0.180	0.0674	0.122	0.143	0.202	0.0259	0.167
Kleibergen-Paap F-statistic	81.24	25.13	68.91	59.07	85.55	73.85	52.38	76.55	54.16	79.46	78.55	50.98	21.96	74.50
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	811.4	46.51	1091.3	249.6	559.5	208.3	332.6	576.7	211.6	427.3	271.2	112.7	116.3	1027.9
Windmeijer F-statistic	17.52	11.09	14.60	7.04	14.48	16.04	6.00	15.97	2.30	10.94	10.78	6.37	10.46	9.10
Kleibergen-Paap Wald F-statistic	19.18	4.987	15.81	13.23	20.38	17.55	11.37	18.18	11.46	18.62	18.63	11.27	4.448	17.62
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 * *p < .05 * * * p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-18: Heterogeneous Effects of the Complementary of Internet and Roads on Agricultural jobs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Farming is main employment							
	All Sample	Urban Sample	Rural Sample	Male Sample	Female Sample	Not educated Sample	Low educated Sample (Primary)	Not more than primary Sample	higher educated Sample	20-34 groupe age Sample	35 -44 groupe age Sample	45-64 groupe age Sample	More developed Sample	Less developed Sample
Submarine internet	-0.112***	0.118	-0.128***	-0.0781**	-0.129***	-0.262***	-0.109***	-0.196***	0.0413**	-0.115***	-0.121***	-0.0828*	-0.183*	-0.102***
	(0.0226)	(0.0673)	(0.0278)	(0.0276)	(0.0245)	(0.0611)	(0.0261)	(0.0347)	(0.0139)	(0.0242)	(0.0254)	(0.0325)	(0.0791)	(0.0228)
Proximity to road (KM)	0.00270***	-0.00460***	0.00601***	-0.000714	0.00404***	0.00184***	0.00430***	0.00307***	-0.000891	0.00254***	0.00293***	0.00275***	0.00738***	0.00165***
	(0.000396)	(0.000877)	(0.000584)	(0.000574)	(0.000447)	(0.000397)	(0.000905)	(0.000413)	(0.000622)	(0.000410)	(0.000462)	(0.000570)	(0.00114)	(0.000427)
Submarine internet * road	-0.0263***	0.0583*	-0.0169***	-0.0251***	-0.0275***	-0.0389***	-0.0236***	-0.0331***	0.00658	-0.0283***	-0.0254***	-0.0217***	-0.0617**	-0.0152***
	(0.00495)	(0.0290)	(0.00350)	(0.00643)	(0.00516)	(0.0101)	(0.00494)	(0.00618)	(0.00397)	(0.00549)	(0.00517)	(0.00661)	(0.0190)	(0.00453)
N	752131	286213	465918	244272	507859	295085	195580	490665	261466	448254	201003	102874	292000	460131
ymean	0.284	0.071	0.415	0.375	0.240	0.403	0.336	0.377	0.111	0.248	0.314	0.382	0.246	0.308
r2	0.220	-0.373	0.165	0.311	0.181	0.222	0.180	0.188	0.150	0.193	0.219	0.267	-0.111	0.275
r2_a	0.220	-0.373	0.164	0.311	0.181	0.221	0.180	0.188	0.150	0.193	0.218	0.267	-0.111	0.275
Kleibergen-Paap F-statistic	80.89	23.87	70.07	58.19	85.11	74.85	52.47	76.84	54.43	78.63	77.85	51.02	21.37	74.14
$P > F$	0	9.12e-259	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	885.1	45.01	1219.3	288.8	596.1	247.4	347.0	632.1	229.5	471.4	289.5	121.8	134.8	1043.2
Windmeijer F-statistic	35.21	10.40	20.23	15.99	24.92	9.81	17.97	12.01	13.24	11.08	10.80	10.42	14.58	9.19
Kleibergen-Paap Wald F-statistic	18.99	4.748	16.04	13.04	20.12	17.78	11.34	18.10	11.56	18.33	18.39	11.24	4.320	17.49
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-19: Heterogeneous Effects of the Complementary of Internet and Roads on Manual Skilled jobs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Manual skilled							
	All Sample	**Urban Sample**	**Rural Sample**	**Male Sample**	**Female Sample**	**Not educated Sample**	**Low educated Sample (Primary)**	**Not more than primary Sample**	**higher educated Sample**	**20-34 groupe age Sample**	**35 -44 groupe age Sample**	**45-64 groupe age Sample**	**More developed Sample**	**Less developed Sample**
Submarine internet	0.0408***	0.0680*	0.0458***	0.123***	0.000656	0.0626**	0.0487***	0.0615***	0.0227	0.0309**	0.0497***	0.0592**	0.0386	0.0374***
	(0.00927)	(0.0337)	(0.0111)	(0.0244)	(0.00655)	(0.0212)	(0.0129)	(0.0127)	(0.0129)	(0.0102)	(0.0122)	(0.0185)	(0.0221)	(0.00880)
Proximity to road (KM)	0.000688***	0.00143***	0.00000773	0.00236***	0.0000219	0.000209	0.00169***	0.000435**	0.00220***	0.000656***	0.000900***	0.000445	0.000808*	0.000638***
	(0.000146)	(0.000433)	(0.000159)	(0.000396)	(0.000117)	(0.000134)	(0.000375)	(0.000139)	(0.000507)	(0.000164)	(0.000206)	(0.000310)	(0.000366)	(0.000155)
Submarine internet * road	0.00850***	0.0273	0.00470**	0.0244***	0.000479	0.00870*	0.00835***	0.00919***	0.00691	0.00719**	0.00907***	0.0117**	0.0111*	0.00564**
	(0.00203)	(0.0144)	(0.00145)	(0.00566)	(0.00131)	(0.00349)	(0.00237)	(0.00225)	(0.00363)	(0.00228)	(0.00246)	(0.00374)	(0.00530)	(0.00172)
N	706524	271241	435283	186423	520101	280275	183819	464094	242430	426452	188634	91438	268532	437992
ymean	0.082	0.116	0.062	0.188	0.044	0.054	0.093	0.070	0.106	0.083	0.080	0.085	0.094	0.075
r2	0.0803	0.0653	0.0572	0.192	0.0324	0.0286	0.100	0.0611	0.108	0.0872	0.0801	0.0722	0.0572	0.0951
r2_a	0.0802	0.0651	0.0571	0.192	0.0323	0.0284	0.100	0.0610	0.108	0.0871	0.0798	0.0717	0.0571	0.0950
Kleibergen-Paap F-statistic	81.39	25.12	69.06	59.96	85.73	73.75	53.23	77.02	53.92	79.65	78.36	51.11	22.52	74.13
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	798.3	46.52	1056.8	237.6	559.2	204.8	326.2	564.8	209.5	422.5	265.9	109.2	116.4	1010.0
Windmeijer F-statistic	18.37	17.32	17.77	13.68	12.03	7.20	7.80	8.28	16.54	9.15	13.78	7.18	8.28	6.71
Kleibergen-Paap Wald F-statistic	19.18	4.985	15.81	13.32	20.42	17.51	11.57	18.26	11.39	18.64	18.52	11.29	4.560	17.52
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

Table 3-20: Heterogeneous Effects of the Complementary of Internet and Roads on Services jobs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
							Services							
	All Sample	**Urban Sample**	**Rural Sample**	**Male Sample**	**Female Sample**	**Not educated Sample**	**Low educated Sample (Primary)**	**Not more than primary Sample**	**higher educated Sample**	**20-34 groupe age Sample**	**35 -44 groupe age Sample**	**45-64 groupe age Sample**	**More developed Sample**	**Less developed Sample**
Submarine internet	0.168***	0.195**	0.121***	0.207***	0.156***	0.235***	0.141***	0.198***	0.0979***	0.177***	0.155***	0.144***	0.273**	0.121***
	(0.0236)	(0.0711)	(0.0207)	(0.0327)	(0.0227)	(0.0490)	(0.0245)	(0.0288)	(0.0259)	(0.0282)	(0.0242)	(0.0314)	(0.0839)	(0.0183)
Proximity to road (KM)	0.00126***	0.00359***	-0.000300	0.00298***	0.000664*	0.000426	0.00161**	0.000718**	0.00424***	0.00140***	0.00139***	0.000627	0.000622	0.00131***
	(0.000283)	(0.000862)	(0.000264)	(0.000522)	(0.000266)	(0.000235)	(0.000554)	(0.000236)	(0.000969)	(0.000316)	(0.000342)	(0.000505)	(0.000806)	(0.000284)
Submarine internet * road	0.0294***	0.0627*	0.0124***	0.0348***	0.0268***	0.0331***	0.0217***	0.0284***	0.0198**	0.0331***	0.0233***	0.0262***	0.0603**	0.0172***
	(0.00529)	(0.0307)	(0.00272)	(0.00772)	(0.00487)	(0.00815)	(0.00467)	(0.00521)	(0.00763)	(0.00651)	(0.00493)	(0.00655)	(0.0206)	(0.00370)
N	717891	278417	439474	197714	520177	281217	185596	466813	251078	431907	191991	93993	274836	443055
ymean	0.226	0.356	0.144	0.401	0.160	0.097	0.197	0.136	0.397	0.221	0.230	0.241	0.267	0.202
r2	0.0956	-0.00814	0.0935	0.205	0.0137	-0.0978	0.0789	0.00246	0.113	0.0595	0.156	0.166	-0.219	0.156
r2_a	0.0955	-0.00832	0.0934	0.204	0.0136	-0.0979	0.0787	0.00235	0.113	0.0594	0.156	0.166	-0.219	0.156
Kleibergen-Paap F-statistic	81.55	25.20	69.03	59.68	85.73	73.73	53.06	76.77	55.55	79.75	78.44	51.45	22.63	74.05
$P > F$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cragg-Donald Wald F-statistic	808.3	47.13	1065.5	246.9	559.2	205.3	327.8	567.1	216.4	426.7	269.3	111.8	118.4	1018.2
Windmeijer F-statistic	20.01	11.52	12.23	14.42	12.87	8.09	18.85	19.42	10.79	19.79	12.87	10.15	10.92	11.44
Kleibergen-Paap Wald F-statistic	19.23	5.006	15.81	13.30	20.42	17.50	11.52	18.19	11.77	18.69	18.52	11.38	4.581	17.48
Age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓
Gender	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place of residence	✓	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grid (20*20) connectivity FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses; clustered at the grid level. $*p < .10 **p < .05 ***p < 0.01$. *We use a strategy that combines difference-in-differences and instrumental variables techniques. The causal effect of Internet is estimated using difference-in-differences, while the effects of the road and its interaction with the Internet are estimated using an instrumental variable approach. Our instruments are: the distance to the hypothetical road, the interaction between the distance to the hypothetical road and an indicator for whether an individual (or a firm) is connected to submarine Internet, lightning, and terrain elevation.*

References

Abbasi, M., Lebrand, M. S. M., Mongoue, A. B., Pongou, R., & Zhang, F. (2022, March). *Roads, Electricity, and Jobs: Evidence of Infrastructure Complementarity in Sub-Saharan Africa* (Policy Research Working Paper Series No. 9976). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/9976.html

Abbasi, M., & Pongou, R. (2022). Aids in the digital age: Evidence from submarine internet in africa.

Agenor, P.-R., Aizenman, J., & Hoffmaister, A. W. (2004, February). The credit crunch in East Asia: what can bank excess liquid assets tell us? *Journal of International Money and Finance*, *23*(1), 27-49. Retrieved from https://ideas.repec.org/a/eee/jimfin/v23y2004i1p27-49.html

Allen, F., & Giovannetti, G. (2011). The effects of the financial crisis on sub-saharan africa. *Review of Development Finance*, *1*(1), 1-27. Retrieved from https://www.sciencedirect.com/science/article/pii/S1879933710000035 doi: https://doi.org/10.1016/j.rdf.2010.10.002

Allen, T. (2014). Information frictions in trade. *Econometrica*, *82*(6), 2041–2083. Retrieved 2022-04-30, from http://www.jstor.org/stable/43616907

Allen N., B., Leora F., K., & Gregory F., U. (2001). The ability of banks to lend to informationally opaque small businesses. *Journal of Banking Finance*, *25*(12), 2127-2167. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378426601001893 doi: https://doi.org/10.1016/S0378-4266(01)00189-3

Almeida, R. K., Carlos H. L, C., & Poole, J. (2017). *The Impact of Digital Technologies on Worker Tasks: Do Labor Policies Matter?* (Tech. Rep.).

Almond, D., & Edlund, L. (2007, 10). Trivers-willard at birth and one year: Evidence from us natality data 1983-2001. *Proceedings. Biological sciences / The Royal Society*, *274*, 2491-6. doi: 10.1098/rspb.2007.0524

Altindag, O. (2016, 03). Son preference, fertility decline, and the nonmissing girls of turkey. *Demography*, *53*. doi: 10.1007/s13524-016-0455-0

Andersen, T., & Hviid, M. (1990). imperfect competition, information acquisition, and price adjustment. In N. Christodoulakis (Ed.), *Dynamic modelling and control of national economies 1989* (p. 7-11). Amsterdam: Pergamon. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780080375380500071 doi: https://doi.org/10.1016/B978-0-08-037538-0.50007-1

Anderson, S. (2018, June). Legal origins and female hiv. *American Economic Review*, *108*(6), 1407-39. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.20151047 doi: 10.1257/aer.20151047

Anderson, S., & Ray, D. (2010). Missing women: Age and disease. *The Review of Economic Studies*, *77*(4), 1262–1300. Retrieved from http://www.jstor.org/stable/40836647

Anderson, S., & Ray, D. (2018, 07). Missing Unmarried Women. *Journal of the European Economic Association*, *17*(5), 1585-1616. Retrieved from https://doi.org/10.1093/jeea/jvy027 doi: 10.1093/jeea/jvy027

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

Arnold, & Beckman, M. (2013, 11). The informed brain in a digital world[U+202F]: interdisciplinary team summaries[U+202F]. *National Academies Press.*, 15-17.

Arnold, F. (1997). *Gender preferences for children.* DHS Comparative Studies. Retrieved from http://dhsprogram.com/pubs/pdf/CS23/CS23.pdf

Aschauer, D. A. (1989). Is public expenditure productive? [Journal Article Report, Technical report, Mimeo .]. (23582).

Atack, J., Bateman, F., Haines, M., & Margo, R. (2009, 01). Did railroads induce or follow economic growth? urbanization and population growth in the american midwest, 1850-60. *Boston University - Department of Economics, Boston University - Department of Economics - The Institute for Economic Development Working Papers Series*, *34*. doi: 10.2307/40587344

Ayyagari, M., Beck, T., & Demirguc-Kunt, A. (2007). Small and medium enterprises across the globe. *Small Business Economics*, *29*(4), 415–434. Retrieved 2022-07-05, from http://www.jstor.org/stable/40229581

Ayyagari, M., Demirguc-Kunt, A., & Maksimovic, V. (2011, April). *Small vs. young firms across the world : contribution to employment, job creation, and growth* (Policy Research Working Paper Series No. 5631). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/5631.html

Bahia, K., Castells, P., Cruz, G., Masaki, T., Pedros, X., Pfutze, T., . . . Winkler, H. J. (2020, May). *The Welfare Effects of Mobile Broadband Internet : Evidence from Nigeria* (Policy Research Working Paper Series No. 9230). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/9230.html

Bahia, K., Castells, P., Cruz, G., Masaki, T., Rodriguez Castelan, C., & Sanfelice, V. (2021, September). *Mobile Broadband Internet, Poverty and Labor Outcomes in Tanzania* (IZA Discussion Papers No. 14720). Institute of Labor Economics (IZA). Retrieved from https://ideas.repec.org/p/iza/izadps/dp14720.html

Balling, M., Bernet, B., Gnan, E., Ayadi, R., Westerfeld, S., Franck, T., . . . Veugelers, R. (2009). *Financing smes in europe.* Citeseer.

Barbosa, E. G., & Moraes, C. D. C. (2003, February). *Determinants Of The Firm's Capital Structure - The Case Of The Very Small Enterprises* (Finance No. 0302001). University Library of Munich, Germany. Retrieved from https://ideas.repec.org/p/wpa/wuwpfi/0302001.html

Barro, R. J. (1976, November). The Loan Market, Collateral, and Rates of Interest. *Journal of Money, Credit and Banking*, *8*(4), 439-456. Retrieved from https://ideas.repec.org/a/mcb/jmoncb/v8y1976i4p439-56.html

Basu, D., & Jong, R. (2010, 05). Son targeting fertility behavior: Some consequences and determinants. *Demography*, *47*, 521-36. doi: 10.1353/dem.0.0110

Bathaluddin, M. B., P, N. M. A., & A.W., W. (2012, January). The impact Of excess liquidity on monetary policy. *Bulletin of Monetary Economics and Banking*, *14*(3), 1-23. Retrieved from https://ideas.repec.org/a/idn/journl/v14y2012i3fp1-23.html doi: https://doi.org/10.21098/

Batjargal, B. (2007). Internet entrepreneurship: Social capital, human capital, and performance of internet ventures in china. *Research Policy*, *36*(5), 605-618. Retrieved from https://www.sciencedirect.com/science/article/pii/S0048733307000686 doi: https://doi.org/10.1016/j.respol.2006.09.029

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, *99*(3), 435–448.

Beck, T. (2013, August). Bank Financing for SMEs - Lessons from the Literature. *National Institute Economic Review*, *225*(1), 23-38. Retrieved from https://ideas.repec.org/a/sae/niesru/v225y2013i1pr23-r38.html

Beck, T., & Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking Finance*, *30*(11), 2931-2943. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378426606000926 doi: https://doi.org/10.1016/j.jbankfin.2006.05.009

Beck, T., Demirguc-Kunt, A., & Martinez Peria, M. (2011, 09). Bank financing for smes: Evidence across countries and bank-ownership types. *Journal of Financial Services Research*, *39*, 35-54. doi: 10.1007/s10693-010-0085-4

Bedard, K., & Deschenes, O. (2003, 01). Sex preferences, marital dissolution and the economic status of women. *Department of Economics, UC Santa Barbara, University of California at Santa Barbara, Economics Working Paper Series*, *40*. doi: 10.2307/4129531

Bencivenga, V. R., & Smith, B. D. (1991, 04). Financial Intermediation and Endogenous Growth. *The Review of Economic Studies*, *58*(2), 195-209. Retrieved from https://doi.org/10.2307/2297964 doi: 10.2307/2297964

Benjamin, D. K. (1978, July). The Use of Collateral to Enforce Debt Contracts. *Economic Inquiry*, *16*(3), 333-359. Retrieved from https://ideas.repec.org/a/oup/ecinqu/v16y1978i3p333-59.html

Berger, A. N., & Udell, G. F. (1998). *The economics of small business finance: the roles of private equity and debt markets in the financial growth cycle* (Tech. Rep.).

Berger., A. N., & Udell., G. F. (2005, December). *A more complete conceptual framework for financing of small and medium enterprises* (Policy Research Working Paper Series No. 3795). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/3795.html

Bertocchia, G., & Dimico, A. (2019, 05). The long-term determinants of female hiv infection in africa: The slave trade, polygyny, and sexual behavior. *Journal of Development Economics*, *140*. doi: 10.1016/j.jdeveco.2019.05.005

Bertschek, I., Polder, M., & Schulte, P. (2019, 01). Ict and resilience in times of crisis: evidence from cross-country micro moments data. *Economics of Innovation and New Technology*, *28*, 1-16. doi: 10.1080/10438599.2018.1557417

Bessone, P., & Dahis, R. (2020). The impact of 3g mobile internet on educational outcomes in brazil∗..

Bhalotra, S. R., & Cochrane, T. (2010, December). *Where Have All the Young Girls Gone? Identification of Sex Selection in India* (IZA Discussion Papers No. 5381). Institute of Labor Economics (IZA). Retrieved from https://ideas.repec.org/p/iza/izadps/dp5381.html

Bhuller, M., Havnes, T., McCauley, J., & Mogstad, M. (2020, November). *How the Internet Changed the Market for Print Media* (Memorandum No. 2/2020). Oslo University, Department of Economics. Retrieved from https://ideas.repec.org/p/hhs/osloec/2020_002.html

Bianchi, N., Lu, Y., & Song, H. (2020, December). *The Effect of Computer-Assisted Learning on Students' Long-Term Development* (NBER Working Papers No. 28180). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/28180.html

Birdsall, N. (1977). Analytical approaches to the relationship of population growth and development. *Population and Development Review*, *3*(1/2), 63–102. Retrieved 2022-04-23, from http://www.jstor.org/stable/1971760

Bonfatti, R., Gu, Y., & Poelhekke, S. (2019). *Priority roads: the political economy of africa's interior-to-coast roads* (Tinbergen Institute Discussion Paper No. TI 2019-006/VIII). Amsterdam and Rotterdam.

Bostrom, N., & Sandberg, A. (2009, 07). Cognitive enhancement: Methods, ethics, regulatory challenges. *Science and engineering ethics*, *15*, 311-41. doi: 10.1007/s11948-009-9142-5

Bozeman, B., & Rogers, J. D. (2002, July). A churn model of scientific knowledge value: Internet researchers as a knowledge value collective. *Research Policy*, *31*(5), 769-794. Retrieved from https://ideas.repec.org/a/eee/respol/v31y2002i5p769-794.html

Brodeur, A., Mabeu, M. C., & Pongou, R. (2020). Ancestral norms, legal origins, and female empowerment. *Working paper*.

Brossard, O., Bocquet, R., & Sabatier, M. (2007, 02). Complementarities in organizational design and the diffusion of information technologies: An empirical analysis. *Research Policy*, *36*, 367-386. doi: 10.1016/j.respol.2006.12.005

Brush, S. G. (1978). Nettie m. stevens and the discovery of sex determination by chromosomes. *Isis*, *69*(2), 163–172. Retrieved from http://www.jstor.org/stable/230427

Buhaug, H., & Rod, J. K. (2006, 03). Local determinants of african civil wars, 1970–2001. *Political geography*, 315-335.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., & i Miquel, G. P. (2015). The value of democracy: Evidence from road building in kenya. *The American Economic Review*, *105*(6), 1817–1851.

Butcher, K. F., & Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, *109*(3), 531–563. Retrieved from https://doi.org/10.2307/2118413

Cagnacci, A., Renzi, A., Arangino, S., Alessandrini, C., & Volpe, A. (2004, 02). Influences of maternal weight on the secondary sex ratio of human offspring. *Human reproduction (Oxford, England)*, *19*, 442-4. doi: 10.1093/humrep/deh071

Campbell, E., & Campbell, P. (1997, 05). Family size and sex preferences and eventual fertility in botswana. *Journal of biosocial science*, *29*, 191-204. doi: 10.1017/S0021932097001910

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

Carrera, J. (2011, 07). Obstetric ultrasounds in africa: Is it necessary to promote their appropriate use? *Donald School Journal of Ultrasound in Obstetrics and Gynecology*, *5*, 289-296. doi: 10.5005/jp-journals-10009-1205

Chen, S., Liu, W., & Song, H. (2020, July). Broadband Internet, Firm Performance, And Worker Welfare: Evidence And Mechanism. *Economic Inquiry*, *58*(3), 1146-1166. Retrieved from https://ideas.repec.org/a/bla/ecinqu/v58y2020i3p1146-1166.html doi: 10.1111/ecin.12854

Choi, E. J., & Hwang, J. (2015, May). Child gender and parental inputs: No more son preference in korea? *American Economic Review*, *105*(5), 638-43. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.p20151118 doi: 10.1257/aer.p20151118

Christensen, C. A. P. . A. M. G. . K. J. C. . S. W. . J. E. (2006, August). *Sub-Saharan Africa : Financial Sector Challenges* (Book No. 9781589065659). International Monetary Fund. Retrieved from https://www.imf.org/en/Publications/Books/Issues/2016/12/31/Sub-Saharan-Africa-Financial-Sector-Challenges-19293

Chun, N., & Tang, H. (2018, May). *Do Information and Communication Technologies Empower Female Workers? Firm-Level Evidence from Viet Nam* (Working Papers No. id:12804). eSocialSciences. Retrieved from https://ideas.repec.org/p/ess/wpaper/id12804.html

Clark, S. (2000). Son preference and sex composition of children: Evidence from india. *Demography*, *37*(1), 95–108. Retrieved from http://www.jstor.org/stable/2648099

Cockburn, I. M., & Wagner, S. (2007, June). *Patents and the survival of internet-related ipos* (Working Paper No. 13146). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13146 doi: 10.3386/w13146

Coco, G. (2000, 02). On the use of collateral. *Journal of Economic Surveys*, *14*, 191-214. doi: 10.1111/1467-6419.00109

Couture, V., Faber, B., Gu, Y., & Liu, L. (2021, March). Connecting the Countryside via E-Commerce: Evidence from China. *American Economic Review: Insights*, *3*(1), 35-50. Retrieved from https://ideas.repec.org/a/aea/aerins/v3y2021i1p35-50.html doi: 10.1257/aeri.20190382

Crews, D. (1996, 03). Temperature-dependent sex determination: The interplay of steroid hormones and temperature. *Zoological science*, *13*, 1-13. doi: 10.2108/zsj.13.1

Cumber, S., & Nchanji, K. (2017, 12). Diagnostic medical ultrasound awareness among women in sub-saharan africa (ssa). *International Journal of Radiology Radiation Therapy*, *4*. doi: 10.15406/ijrrt.2017.04.00120

Cusolito, A. P., Safadi, R., & Taglioni, D. (2016). *Inclusive global value chains: Policy options for small and medium enterprises and low-income countries.* World Bank Publications.

Dahl, G. B., & Moretti, E. (2004, February). *The Demand for Sons: Evidence from Divorce, Fertility, and Shotgun Marriage* (NBER Working Papers No. 10281). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/10281.html

Das Gupta, M., & Bhat, P. (1997, 11). Fertility decline and increased manifestation of sex bias in india. *Population Studies*, *51*, 307-315. doi: 10.1080/0032472031000150076

Demetriades, P. O., & Hussein, K. A. (1996). Does financial development cause economic growth? time-series evidence from 16 countries. *Journal of Development Economics*, *51*(2), 387-411. Retrieved from https://www.sciencedirect.com/science/article/pii/S030438789600421X doi: https://doi.org/10.1016/S0304-3878(96)00421-X

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

Derksen, L., Leclerc, C. M., & Souza, P. C. (2019). *Searching for Answers : The Impact of Student Access to Wikipedia* (Tech. Rep.).

Desgranges, G. (2000, August). *CK-Equilibria and Informational Efficiency in a Competitive Economy* (Econometric Society World Congress 2000 Contributed Papers No. 1296). Econometric Society. Retrieved from https://ideas.repec.org/p/ecm/wc2000/1296.html

di Mauro, F., Ottaviano, G. I., & Hassan, F. (2017, February). *Banks credit and productivity growth* (Working Paper Series No. 2008). European Central Bank. Retrieved from https://ideas.repec.org/p/ecb/ecbwps/20172008.html

Donaldson, D. (2010, April). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899-934. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20101199 doi: 10.1257/aer.20101199

Draca, M., Sadun, R., & Reenen, J. V. (2006, August). *Productivity and ICT: A Review of the Evidence* (CEP Discussion Papers No. dp0749). Centre for Economic Performance, LSE. Retrieved from https://ideas.repec.org/p/cep/cepdps/dp0749.html

Duarte, J., & Young, L. (2009). Why is pin priced? *Journal of Financial Economics*, *91*(2), 119-138. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304405X08001827 doi: https://doi.org/10.1016/j.jfineco.2007.10.008

Duflo, E., Dupas, P., & Kremer, M. (2015, September). Education, hiv, and early fertility: Experimental evidence from kenya. *American Economic Review*, *105*(9), 2757-97. Retrieved from http://www.aeaweb.org/articles?id=10.1257/aer.20121607 doi: 10.1257/aer.20121607

Dupas, P. (2005, October). *Relative Risks and the Market for Sex: Teenagers, Sugar Daddies and HIV in Kenya* (MPRA Paper No. 248). University Library of Munich, Germany. Retrieved from https://ideas.repec.org/p/pra/mprapa/248.html

Dutz, M. A., Mation, L. F., O'Connell, S. D., & Willig, R. D. (2017). Economy-wide and sectoral impacts on workers of brazil's internet rollout. *Forum for Social Economics*, *46*(2), 160-177. Retrieved from https://doi.org/10.1080/07360932.2017.1307137 doi: 10.1080/07360932.2017.1307137

Ejaz Ghani, W. R. K., Arti Grover Goswami. (2016, March). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing, [Journal Article Report, Technical report, Mimeo .]. (23582).

Elali, W., Gitman, L., Zutter, C., & Roubaie, A. (2013). *Principles of managerial finance.*

Espinoza, H., Kling, G., McGroarty, F., O'Mahony, M., & Ziouvelou, X. (2020). Estimating the impact of the internet of things on productivity in europe. *Heliyon*, *6*(5), e03935. Retrieved from https://www.sciencedirect.com/science/article/pii/S2405844020307805 doi: https://doi.org/10.1016/j.heliyon.2020.e03935

Estache, A., Ianchovichina, E., Bacon, R., & Salamon, I. (2013, 01). Short-run employment effects of infrastructure investment. In (p. 29-59). doi: 10.1596/9780821396650_CH03

Ettredge, M., Richardson, V., & Scholz, S. (2002, 12). Dissemination of information for investors at corporate web sites. *Journal of Accounting and Public Policy*, *21*, 357-369.

Fan, J., Tang, L., Zhu, W., & Zou, B. (2018). The alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, *114*, 203-220. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022199618301594 doi: https://doi.org/10.1016/j.jinteco.2018.07.002

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

Fernandes, A. M., Mattoo, A., Nguyen, H., & Schiffbauer, M. (2019). The internet and Chinese exports in the pre-ali baba era. *Journal of Development Economics*, *138*(C), 57-76. Retrieved from https://ideas.repec.org/a/eee/deveco/v138y2019icp57-76.html doi: 10.1016/j.jdeveco.2018.11

Filmer, D., Friedman, J., & Schady, N. (2008, September). *Development, modernization, and son preference in fertility decisions* (Policy Research Working Paper Series No. 4716). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/4716.html

Fjose, C., & Grunfeld, L. (2010). Smes and growth in sub-saharan africa, identifying sme roles and obstacles to sme growth. *MENON Business Economics*, *14/2010*(1). Retrieved from https://www.menon.no/publication/smes-and-growth-in-sub-saharan-africa-identifying-sme-roles-and-obstacles-to-sme-growth/

Friedman, M., & Schwartz, A. J. (1963). *A monetary history of the united states, 1867-1960*. Princeton University Press. Retrieved from http://www.jstor.org/stable/j.ctt7s1vp

Frish, R., & Tsur, S. (2010). Transport Infrastructure Investment, Commuting, and Wages. *Israel Economic Review*, *7*(2), 55-79. Retrieved from https://ideas.repec.org/a/boi/isrerv/v7y2010i2p55-79.html

Gajewski, J.-F., & Li, L. (2015). Can internet-based disclosure reduce information asymmetry? *Advances in Accounting*, *31*(1), 115-124. Retrieved from https://www.sciencedirect.com/science/article/pii/S0882611015000140 doi: https://doi.org/10.1016/j.adiac.2015.03.013

Gakunu, P. (2007). Reforming the financial system in sub-saharan africa: the (long) way ahead. *Finance Bien Commun*, *28-29, no.3*, 139-146.

Galindo, A., & Micco, A. (2016, December). Creditor protection, information sharing and credit for small and medium-sized enterprises: cross-country evidence. *Revista CEPAL*. Retrieved from https://ideas.repec.org/a/ecr/col070/41253.html

Gangadharan, L., & Maitre, P. (1999, 01). Testing for son preference in south africa. *Sydney - Department of Economics, Papers*, *12*. doi: 10.1093/jae/12.3.371

Ganley, J. (2004). *Surplus Liquidity: Implications for Central Banks* (No. 3). Centre for Central Banking Studies, Bank of England. Retrieved from https://ideas.repec.org/b/ccb/lectur/3.html

Garg, A., & Morduch, J. (1998). Sibling rivalry and the gender gap: Evidence from child health outcomes in ghana. *Journal of Population Economics*, *11*(4), 471–493. Retrieved from http://www.jstor.org/stable/20007600

Gelbard, M. E. A., & Leite, M. S. P. (1999, August). *Measuring Financial Development in Sub-Saharan Africa* (IMF Working Papers No. 1999/105). International Monetary Fund. Retrieved from https://ideas.repec.org/p/imf/imfwpa/1999-105.html

Gibbons, S., Lyytikainen, T., Overman, H. G., & Sanchis-Guarner, R. (2016, April). *New road infrastructure: the effects on firms* (CEPR Discussion Papers No. 11239). C.E.P.R. Discussion Papers. Retrieved from https://ideas.repec.org/p/cpr/ceprdp/11239.html

Gigineishvili, N. (2011, 07). Determinants of interest rate pass-through: Do macroeconomic conditions and financial market structure matter? *IMF Working Papers*, *11*. doi: 10.5089/9781462312313.001

Goldfarb, A., & Tucker, C. (2019, March). Digital Economics. *Journal of Economic Literature*, *57*(1), 3-43. Retrieved from https://ideas.repec.org/a/aea/jeclit/v57y2019i1p3-43.html

Gong, F., & Liu, H. (2016). Asymmetric information, heterogeneous prior beliefs, and public information. *International Review of Economics Finance*, *46*, 100-120. Retrieved from https://www.sciencedirect.com/science/article/pii/S1059056016300594 doi: https://doi.org/10.1016/j.iref.2016.07.005

Gordon, R. H., & Li, W. (1995). The change in productivity of chinese state enterprises, 1983–1987. *Journal of Productivity Analysis*, *6*(1), 5–26.

Goyal, A. (2010, May). *Information, direct access to farmers, andrural market performance in central India* (Policy Research Working Paper Series No. 5315). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/5315.html

Grazzi, M., & Jung, J. (2019, 02). What are the drivers of ict diffusion? evidence from latin american firms. *Information Technologies and International Development*, *15*, 34-48.

Greenwood, J., & Jovanovic, B. (1990). *Financial Development, Growth, And The Distribution Of Income* (Tech. Rep.).

Greenwood, J., & Smith, B. D. (1997). Financial markets in development, and the development of financial markets. *Journal of Economic Dynamics and Control*, *21*(1), 145-181. Retrieved from https://www.sciencedirect.com/science/article/pii/0165188995009280 doi: https://doi.org/10.1016/0165-1889(95)00928-0

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393–408. Retrieved from http://www.jstor.org/stable/1805228

Guerrero Barreto, M. E., & Ritter Burga, P. I. (2014). The effect of internet and cell phones on employment and agricultural production in rural villages in peru.

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

Henking, H. (1891). Über spermatogenese und deren beziehung zur entwicklung. *Pyrrhocoris apterus L. Zeitschrift für wissenschaftliche Zoologie*, 685–736.

Herbst, J. (2000, 10). 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., & Lebrand, M. (2021). *Infrastructure and structural change in the horn of africa.*

Hisrich, R., & Drnovsek, M. (2002, 06). Entrepreneurship and small business research–a european perspective. *Journal of Small Business and Enterprise Development*, *9*, 172-222. doi: 10.1108/14626000210427348

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

Hjort, J., & Poulsen, J. (2019b, March). The Arrival of Fast Internet and Employment in Africa. *American Economic Review*, *109*(3), 1032-1079. Retrieved from https://ideas.repec.org/a/aea/aecrev/v109y2019i3p1032-79.html

Hjort, J., & Tian, L. (2021, November). The Economic Impact of Internet Connectivity in Developing Countries . *INSEAD Working Paper, SSRN*. Retrieved from https://ideas.repec.org/a/aea/jeclit/v57y2019i1p3-43.html

Hodge, F., Jollineau, J., & Maines, L. (2003, 06). Does search[U+2010]facilitating technology improve the transparency of financial reporting? *The Accounting Review*, *79*. doi: 10.2308/accr.2004.79.3.687

Holl, A. (2004, November). Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal. *Journal of Regional Science*, *44*(4), 693-712. Retrieved from https://ideas.repec.org/a/bla/jregsc/v44y2004i4p693-712.html doi: 10.1111/j.0022-4146.2004.

Houngbonon, G. V., Mensah, J. T., & Traore, N.-. (2022, February). *The Impact of Internet Access on Innovation and Entrepreneurship in Africa* (Policy Research Working Paper Series No. 9945). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/9945.html

Hu, L., & Schlosser, A. (2015, September). Prenatal Sex Selection and Girls' Well[U+2010]Being: Evidence from India. *Economic Journal*, *125*(587), 1227-1261. Retrieved from https://ideas.repec.org/a/wly/econjl/v125y2015i587p1227-1261.html

Idrees, R. Q., Baig, M. S. R., & Nawaz, H. M. U. (2020). Export oriented smes of pakistan in perspective of cpec. *Pakistan Journal of International Affairs*, *3*(2).

Ireland, P. N. (1994, March). Money and Growth: An Alternative Approach. *American Economic Review*, *84*(1), 47-65. Retrieved from https://ideas.repec.org/a/aea/aecrev/v84y1994i1p47-65.html

Irwin, D., & Scott, J. (2010, 05). Barriers faced by smes in raising bank finance. *International Journal of Entrepreneurial Behaviour amp Research*, *16*, 245-259. doi: 10.1108/13552551011042816

James, W. (1995, 05). What stabilizes the sex ratio? *Annals of human genetics*, *59*, 243-9. doi: 10.1111/j.1469-1809.1995.tb00744.x

James, W. (1998, 06). Hypothesis: One cause of polydactyly. *Journal of theoretical biology*, *192*, 1-2. doi: 10.1006/jtbi.1997.0591

James, W. H. (1996). Evidence that mammalian sex ratios at birth are partially controlled by parental hormone levels at the time of conception. *Journal of Theoretical Biology*, *180*(4), 163–172. Retrieved from https://doi.org/10.1006/jtbi.1996.0102

James M Cook, R. H. C. (1995, july). Sex determination and population biology in the hymenoptera. *Trends in Ecology Evolution*, *270*, 71 - 78. Retrieved from https://doi.org/10.1016/0169-5347(95)90011-X

Jayachandran, S. (2017a, January). Fertility Decline and Missing Women. *American Economic Journal: Applied Economics*, *9*(1), 118-139. Retrieved from https://ideas.repec.org/a/aea/aejapp/v9y2017i1p118-39.html

Jayachandran, S. (2017b). Odds are you're measuring son preference incorrectly. *Development Impact*. Retrieved from https://blogs.worldbank.org/impactevaluations/odds-are-you-re-measuring-son-preference-incorrectly

Jayachandran, S., & Kuziemko, I. (2011). Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India. *The Quarterly Journal of Economics*, *126*(3), 1485-1538. Retrieved from https://ideas.repec.org/a/oup/qjecon/v126y2011i3p1485-1538.html

Jayachandran, S., & Pande, R. (2017, September). Why Are Indian Children So Short? The Role of Birth Order and Son Preference. *American Economic Review*, *107*(9), 2600-2629. Retrieved from https://ideas.repec.org/a/aea/aecrev/v107y2017i9p2600-2629.html

Jean-Pierre Briffaut, M. P. C. H. M. T. L. M. H. M. M. P. U., Mr. George Iden, & Swaray, M. S. M. (2018). Broadening financial intermediation in sub-saharan africa", booktitle=financing sustainable development in africa. In (pp. 93–119). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-78843-2_5 doi: 10.1007/978-3-319-78843-2_5

Jedwab, R., & Storeygard, A. (2017, August). *The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960-2010* (Working Paper No. 27670). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w27670 doi: 10.3386/w27670

Jedwab, R., & Storeygard, A. (2021, 06). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960-2010. *Journal of the European Economic Association*.

Jovanovic, B., Krstevska, A., & Popovska-Kamnar, N. (2015, October). *Can Monetary Policy Affect Economic Activity under Surplus Liquidity? Some Evidence from Macedonia* (Working Papers No. 2015-03). National Bank of the Republic of North Macedonia. Retrieved from https://ideas.repec.org/p/mae/wpaper/2015-03.html

Juhn, C., Ujhelyi, G., & Villegas-Sanchez, C. (2013, May). Trade Liberalization and Gender Inequality. *American Economic Review*, *103*(3), 269-273. Retrieved from https://ideas.repec.org/a/aea/aecrev/v103y2013i3p269-73.html

Juntunen, K., Kvist, A.-P., & Kauppila, A. (1997, 10). A shift from a male to a female majority in newborns with the increasing age of grand grand multiparous women. *Human reproduction (Oxford, England)*, *12*, 2321-3. doi: 10.1093/humrep/12.10.2321

Kazimoto, P. (2014, April). Assessment of Challenges facing Small and Medium Enterprises towards International Marketing Standards: a Case Study of Arusha Region Tanzania. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *4*(2), 303-311. Retrieved from https://ideas.repec.org/a/hur/ijaraf/v4y2014i2p303-311.html

Khalifa Al-Yousif, Y. (2002). Financial development and economic growth: Another look at the evidence from developing countries. *Review of Financial Economics*, *11*(2), 131-150. Retrieved from https://www.sciencedirect.com/science/article/pii/S1058330002000393 doi: https://doi.org/10.1016/S1058-3300(02)00039-3

Khanna, R., & Sharma, C. (2018). Testing the effect of investments in it and rd on labour productivity: New method and evidence for indian firms. *Economics Letters*, *173*, 30-34. Retrieved from https://www.sciencedirect.com/science/article/pii/S0165176518303811 doi: https://doi.org/10.1016/j.econlet.2018.09.003

Khemraj, T. (2008, February). *Excess liquidity, oligopolistic loan markets and monetary policy in LDCs* (Working Papers No. 64). United Nations, Department of Economics and Social Affairs. Retrieved from https://ideas.repec.org/p/une/wpaper/64.html

Kho, K., Lakdawala, L., & Nakasone, E. (2018, April). *Impact of Internet Access on Student Learning in Peruvian Schools* (Working Papers No. 2018-3). Michigan State University, Department of Economics. Retrieved from https://ideas.repec.org/p/ris/msuecw/2018_003.html

Kihlstrom, R. E., & Mirman, L. J. (1975). Information and market equilibrium. *The Bell Journal of Economics*, *6*(1), 357–376. Retrieved from http://www.jstor.org/stable/3003230

Kira, A. (2013, 03). Determinants of financing constraints in east african countries' smes. *International Journal of Business and Management*, *8*. doi: 10.5539/ijbm.v8n8p49

Kira, A., & He, Z. (2012, 11). The impact of firm characteristics in access of financing by small and medium-sized enterprises in tanzania. *International Journal of Business and Management*, *7*. doi: 10.5539/ijbm.v7n24p108

Kitindi, E., B.A.S, M., & Sethibe, A. (2007, 02). Lending decision making and financial information: The usefulness of corporate annual reports to lenders in botswana. *The International Journal of Applied Economics and Finance*. doi: 10.3923/ijaef.2007.55.66

Kolehmainen, S., Coulthard, M., & Loos, J. (2007). Networking, the missing link in firm based entrepreneurial orientation (eo) performance models. In (pp. 1 – 21). (International Council for Small Business Conference (ICSB) 2007 ; Conference date: 13-06-2007 Through 15-06-2007)

Kremer, M. (1996, January). *Integrating Behavioral Choice into Epidemiological Models of the AIDS Epidemic* (NBER Working Papers No. 5428). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/5428.html

Kugler, A. D., & Kumar, S. (2017, June). Preference for Boys, Family Size, and Educational Attainment in India. *Demography*, *54*(3), 835-859. Retrieved from https://ideas.repec.org/a/spr/demogr/v54y2017i3d10.1007_s13524-017-0575-1.html doi: 10.1007/s13524-017-0575-1

Kuhn, P. J., & Mansour, H. (2011, September). *Is Internet Job Search Still Ineffective?* (IZA Discussion Papers No. 5955). Institute of Labor Economics (IZA). Retrieved from https://ideas.repec.org/p/iza/izadps/dp5955.html

Larsen, U., Chung, W., & Gupta, M. D. (1998). Fertility and son preference in korea. *Population Studies*, *52*(3), 317–325. Retrieved from http://www.jstor.org/stable/2584733

Le, N. T., & Nguyen, T. V. (2009, July). The Impact of Networking on Bank Financing: The Case of Small and Mediumâ€“Sized Enterprises in Vietnam. *Entrepreneurship Theory and Practice*, *33*(4), 867-887. Retrieved from https://ideas.repec.org/a/sae/entthe/v33y2009i4p867-887.html doi: 10.1111/j.1540-6520.2009.

Lebrand, M. (2022, 1). *Infrastructure and structural change in the lake chad region* (Policy Research Working Paper Series No. WPS 9899). The World Bank.

Lederman, D., & Zouaidi, M. (2020, March). *Incidence of the Digital Economy and Frictional Unemployment : International Evidence* (Policy Research Working Paper Series No. 9170). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/9170.html

Leung, S. (1991, 02). A stochastic dynamic analysis of parental sex preferences and fertility. *The Quarterly Journal of Economics*, *106*, 1063-88. doi: 10.2307/2937957

Leuven, E., Akerman, A., & Mogstad, M. (2018, February). *Information Frictions, Internet and the Relationship between Distance and Trade* (Memorandum No. 1/2018). Oslo University, Department of Economics. Retrieved from https://ideas.repec.org/p/hhs/osloec/2018_001.html

Levine, R. (2004, 01). Finance and growth: Theory and evidence. *National Bureau of Economic Research, Inc, NBER Working Papers*, *10766*.

Li, H., & Li, Z. (2013). Road investments and inventory reduction: Firm level evidence from china. *Journal of Urban Economics*, *76*, 43–52.

Li, L., & Wu, X. (2011). Gender of children, bargaining power, and intrahousehold resource allocation in china. *The Journal of Human Resources*, *46*(2), 295–316. Retrieved from http://www.jstor.org/stable/41304821

Li, Z., Wu, M., & Chen, B. R. (2017). Is road infrastructure investment in china excessive? evidence from productivity of firms. *Regional Science and Urban Economics*, *65*, 116-126. Retrieved from https://www.sciencedirect.com/science/article/pii/S0166046217301631 doi: https://doi.org/10.1016/j.regsciurbeco.2017.05.001

Lindahl-Jacobsen, R., Bostofte, E., Engholm, G., Hansen, J., Skakkebaek, N., & Moller, H. (2000, 10). Fertility and offspring sex ratio of men who develop testicular cancer: A record linkage study. *Human reproduction (Oxford, England)*, *15*, 1958-61. doi: 10.1093/humrep/15.9.1958

The long-term effect of demographic shocks on the evolution of gender roles: Evidence from the transatlantic slave trade. (2014). *Journal of the European Economic Association*.

Lundberg, S., & Rose, E. (2003). Child gender and the transition to marriage. *Demography*, *40*(2), 333–349. Retrieved from http://www.jstor.org/stable/3180804

Mabeu, M. C., & Pongou, R. (2019). Democracy, genes, and the male survival disadvantage. *Working paper*.

Mahlknecht, G. (2014, August). *Greg's cable map.* (Working Paper). National Bureau of Economic Research. Retrieved from https://www.cablemap.info/(accessedNovember16,2018). doi: 10.3386/w27670

Mair, L. (1971). *Marriage.* Scolar Press. doi: 085967360X,9780859673600

Malamud, O., Cueto, S., Cristia, J., & Beuermann, D. W. (2019). Do children benefit from internet access? experimental evidence from peru. *Journal of Development Economics*, *138*, 41-56. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387818301251 doi: https://doi.org/10.1016/j.jdeveco.2018.11.005

Mammen, K. (2008, 04). The effect of children's gender on living arrangements and child support. *American Economic Review - AER*, *98*, 408-412. doi: 10.1257/aer.98.2.408

Martin, G., & Staines, H. (1994, October 1). Managerial competences in small firms. *Journal of Management Development*, *13*(7), 23–34. doi: 10.1108/02621719410063396

Masaki, T., Granguillhome Ochoa, R., & Rodriguez Castelan, C. (2020, September). *Broadband Internet and Household Welfare in Senegal* (Policy Research Working Paper Series No. 9386). The World Bank. Retrieved from https://ideas.repec.org/p/wbk/wbrwps/9386.html

Matas, A., Raymond, J.-L., & Roig, J.-L. (2015). Wages and accessibility: The impact of transport infrastructure. *Regional Studies*, *49*(7), 1236-1254. Retrieved from https://doi.org/10.1080/00343404.2013.827336 doi: 10.1080/00343404.2013.827336

Mbekeani, K. K. (2010). Infrastructure, Trade Expansion and Regional Integration: Global Experience and Lessons for Africa-super- †. *Journal of African Economies*, *19*(suppl$_1$), $88 - 113. Retrievedfrom$

McKinnon., R. (1973). Money and capital in economic development. *Brookings Institution, Washington*. doi: hhttps://www.brookings.edu/book/money-and-capital-in-economic-development/

Melo, P. C., Graham, D. J., & Noland, R. B. (2010). Impact of transport infrastructure on firm formation: Evidence from portuguese municipalities. *Transportation Research Record*, *2163*(1), 133-143. Retrieved from https://doi.org/10.3141/2163-15 doi: 10.3141/2163-15

Milazzo, A. (2018). Why are adult women missing? Son preference and maternal survival in India. *Journal of Development Economics*, *134*(C), 467-484. Retrieved from https://ideas.repec.org/a/eee/deveco/v134y2018icp467-484.html doi: 10.1016/j.jdeveco.2018.06

Miller, B., & Atkinson, R. (2014, 01). Raising european productivity growth through ict. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3079844

Mishra, P., Montiel, P. J., & Spilimbergo, A. (2012). Monetary transmission in low-income countries: Effectiveness and policy implications. *IMF Economic Review*, *60*(2), 270–302. Retrieved from http://www.jstor.org/stable/23276895

Mlachila, M., Jidoud, A., Newiak, M., Radzewicz-Bak, B., & Takebe, M. (2016). *Financial development in sub-saharan africa: Promoting inclusive and sustainable growth.* INTERNATIONAL MONETARY FUND. Retrieved from https://books.google.ca/books?id=WpkYEAAAQBAJ

Moller, H., Lindahl-Jacobsen, R., TjØnneland, A., & Overvad, K. (1998, 06). Sex ratio of offspring of diabetics. *Lancet*, *351*, 1514-5. doi: 10.1016/S0140-6736(05)78909-1

Moneke, N. (2019a). *Can big push infrastructure unlock development? evidence from ethiopia* (Tech. Rep.). Technical report, Mimeo.

Moneke, N. (2019b, July). *Can big push infrastructure unlock development? evidence from ethiopia* (Technical Report, Technical report, Mimeo . No. 23582). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23582 doi: 10.3386/w23582

Morduch, J. (2000, 05). Sibling rivalry in africa. *American Economic Review*, *90*, 405-409. doi: 10.1257/aer.90.2.405

Mouelhi, R. B. A. (2009, October). *Impact Of The Adoption Of Information And Communication Technologies On Firm Efficiency In The Tunisian Manufacturing Sector* (Working Papers No. 506). Economic Research Forum. Retrieved from https://ideas.repec.org/p/erg/wpaper/506.html

Mughal, M. Y., & Javed, R. (2018, March). *Have a son, gain a voice: Son preference and female participation in household decision making* (Working Papers No. 2017-2018$_5$)$.CATT-UPPA-UniversitédePauetdesPaysdel'Adour.Retrievedfrom$

Muller-Crepon, C., Hunziker, P., & Cederman, L.-E. (2020, 10). 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., & Deichmann, U. (2004). *The African Population Database. Version 4* (Tech. Rep.).

Nors, T. (1996). Illegitimate children and their high[U+2010]born mothers. *Scandinavian Journal of History*, *21*(1), 17-37. Retrieved from https://doi.org/10.1080/03468759608579313 doi: 10.1080/03468759608579313

Okoye, D., Pongou, R., & Yokossi, T. (2019). New technology, better economy? the heterogeneous impact of colonial railroads in nigeria. *Journal of Development Economics*, *140*, 320–354.

Okpara, J. O. (2011). Factors constraining the growth and survival of smes in nigeria: Implications for poverty alleviation. *Management research review*.

Olawale, F., & Akinwumi, O. (2010, 09). Which new small and medium enterprises in south africa have access to bank credit? *International Journal of Business and Management*, *5*. doi: 10.5539/ijbm.v5n10p128

Osaranen, L. (2008). Africa: Tradition at the heart of violence against women and girls. *Addis Ababa: Inter-African Committee on Traditional Practices (IAC)*.

Oster, E. (2012, 01). Hiv and sexual behavior change: Why not africa? *Journal of health economics*, *31*, 35-49. doi: 10.1016/j.jhealeco.2011.12.006

Owualah, S. (2002, 02). Smes, borrowing constraints and banking relationships in japan. *Japan and the World Economy*, *14*, 87-100. doi: 10.1016/S0922-1425(99)00034-1

Parkhurs, J. O. (2009). Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in african countries. *Bulletin of the World Health Organization*, *88*, 519-526. Retrieved from https://www.who.int/bulletin/volumes/88/7/09-070185/en/

Paterson, A. (1998, 06). Sex ratio of offspring of diabetics. *Lancet*, *351*, 1515. doi: 10.1016/S0140-6736(05)78910-8

POLIQUIN, C. W. (2020). The Wage and Inequality Impacts of Broadband Internet. *UCLA Working paper*.

Pongou, K. D. B., R., & Tsala Dimbuene, Z. (2017). Excess male infant mortality: The gene-institution interactions. *American Economic Review*, *107*(5), 541-45.

Pongou, R. (2008). Is sex imbalance in early age mortality driven by prebirth environmental factors, child biology, or parental preferences? evidence from male-female twin pairs. *Working paper*.

Pongou, R. (2010, April). *A dynamic theory of fidelity networks with an application to the spread of HIV/AIDS* (Working Papers No. 2009-03). Instituto Madrileño de Estudios Avanzados (IMDEA) Ciencias Sociales. Retrieved from https://ideas.repec.org/p/imd/wpaper/wp2009-03.html

Pongou, R. (2013a). Why is infant mortality higher in boys than in girls? a new hypothesis based on preconception environment and evidence from a large sample of twins. *Demography*, *50*(2), 421-444.

Pongou, R. (2013b, 11). Why is infant mortality higher in boys than in girls? a new hypothesis based on preconception environment and evidence from a large sample of twins. *Demography*, *50*. doi: 10.1007/s13524-012-0161-5

Pongou, R. (2020). Is excess (fe)male mortality caused by the prenatal environment, child biology, or parental discrimination? new evidence from male-female twins. *Department of Economics, University of Ottawa. Working Paper 2008E*, *50*(2), 421-444.

Pongou, R., Kuate Defo, B., & Tsala Dimbuene, Z. (2017). Excess male infant mortality: The gene-institution interactions. *American Economic Review*, *107*(5), 541–45.

Pongou, R., & Serrano, R. (2013, May). Fidelity Networks and Long-Run Trends in HIV/AIDS Gender Gaps. *American Economic Review*, *103*(3), 298-302. Retrieved from https://ideas.repec.org/a/aea/aecrev/v103y2013i3p298-302.html

Ponmani, R. (2011). Infrastructure and smes development in selected asian countries. *Asian Journal of Research in Social Sciences and Humanities*, *1*(4), 465–473.

Pontes, G., Teresa, F., & Sol, F. (2021, 06). The determinants of the bank's excess liquidity and the credit crisis: the case of cape verde.

Pretorius, M., & Shaw, G. (2004). Business plans in bank decision-making when financing new ventures in south africa. *South African Journal of Economic and Management Sciences*, *7*, 221-241.

Quartey, P., Turkson, E., Abor, J. Y., & Iddrisu, A. M. (2017). Financing the growth of smes in africa: What are the contraints to sme financing within ecowas? *Review of Development Finance*, *7*(1), 18-28. Retrieved from https://www.sciencedirect.com/science/article/pii/S1879933717300362 (Financial Sector Development in Africa: An Overview) doi: https://doi.org/10.1016/j.rdf.2017.03.001

Rai, P., Paudel, I., Ghimire, A., Pokharel, P., Rijal, R., & Niraula, S. (2014, 02). Effect of gender preference on fertility: Cross-sectional study among women of tharu community from rural area of eastern region of nepal. *Reproductive health*, *11*, 15. doi: 10.1186/1742-4755-11-15

Raleigh, C., & Hegre, H. (2009). Population size, concentration, and civil war. a geographically disaggregated analysis. *Political Geography*, *28*(4), 224-238.

Rao, V., Gupta, I., Lokshin, M., & Jana, S. (2003, August). Sex workers and the cost of safe sex: the compensating differential for condom use among Calcutta prostitutes. *Journal of Development Economics*, *71*(2), 585-603. Retrieved from https://ideas.repec.org/a/eee/deveco/v71y2003i2p585-603.html

Rea, L. M. (1997). *Designing and conducting survey research : a comprehensive guide* (2nd ed. ed.). San Francisco: Jossey-Bass Publishers.

Reardon, T., Berdegué, J., & Escobar, G. (2001). Rural nonfarm employment and incomes in latin america: Overview and policy implications. *World Development*, *29*(3), 395-409. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305750X00001121 (Rural Nonfarm Employment and Incomes in Latin America) doi: https://doi.org/10.1016/S0305-750X(00)00112-1

Reuber, A. R., & Fischer, E. (2011). International entrepreneurship in internet-enabled markets. *Journal of Business Venturing*, *26*(6), 660-679. Retrieved from https://www.sciencedirect.com/science/article/pii/S0883902611000541 doi: https://doi.org/10.1016/j.jbusvent.2011.05.002

Rjasanowski, I., Klöting, I., & Kovacs, P. (1998, 03). Altered sex ratio in offspring of mothers with insulin dependent diabetes. *Lancet*, *351*, 497-8. doi: 10.1016/S0140-6736(05)78685-2

Robinson, J. (1979). The generalisation of the general theory. in: The generalisation of the general theory and other essays. *Palgrave Macmillan, London.*.

Rodrik, D. (2016, March). Premature deindustrialization. *Journal of Economic Growth*, *21*(1), 1-33. Retrieved from https://ideas.repec.org/a/kap/jecgro/v21y2016i1d10.1007_s10887-015-9122-3.html doi: 10.1007/s10887-015-9122-3

Rollo, V., & Paunov, C. (2015, 11). Has the internet fostered inclusive innovation in the developing world? *World Development*, *78*, 587–609. doi: 10.1016/j.worlddev.2015.10.029

Saxegaard, M. (2006, 06). Excess liquidity and the effectiveness of monetary policy: Evidence from sub-saharan africa. *IMF Working Papers (International Monetary Fund)*, *06*. doi: 10.5089/9781451863758.001

Schneebaum, A., & Mader, K. (2013, October). *The gendered nature of intra-household decision making in and across Europe* (Department of Economics Working Papers No. wuwp157). Vienna University of Economics and Business, Department of Economics. Retrieved from https://ideas.repec.org/p/wiw/wiwwuw/wuwp157.html

Seidl, C. (1995, May). The Desire for a Son Is the Father of Many Daughters: A Sex Ratio Paradox. *Journal of Population Economics*, *8*(2), 185-203. Retrieved from https://ideas.repec.org/a/spr/jopoec/v8y1995i2p185-203.html

Sen, A. (1990, 12/20/1990). *More than 100 million women are missing.*

Shane, S., & Cable, D. (2002). Network ties, reputation, and the financing of new ventures. *Management Science*, *48*(3), 364–381. Retrieved from http://www.jstor.org/stable/822571

Shatz, H. J., Kitchens, K. E., Rosenbloom, S., & Wachs, M. (2011). The effects of highway infrastructure on economic activity. In *Highway infrastructure and the economy: Implications for federal policy* (pp. 15–42). RAND Corporation. Retrieved from http://www.jstor.org/stable/10.7249/mg1049rc.9

Shaw, E. S. (1973). Financial deepening in economic development : , (, 1973) xii+260 pp. *Oxford University Press, New York*.

Shen, D., & Chen, S.-H. (2018). Big data finance and financial markets. In S.-H. Chen (Ed.), *Big data in computational social science and humanities* (pp. 235–248). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-95465-3_12 doi: 10.1007/978-3-319-95465-3_12

Smallbone, D., & Welter, F. (2001). The distinctiveness of entrepreneurship in transition economies. *Small Business Economics*, *16*(4), 249–262. Retrieved from http://www.jstor.org/stable/40229156

Song, S., & Burgard, S. A. (2008). Does son preference influence children's growth in height? a comparative study of chinese and filipino children. *Population Studies*, *62*(3), 305–320. Retrieved from http://www.jstor.org/stable/27643477

Southiseng, N., Ty, M., Walsh, J., & Anurit, P. (2008). Development of excellent entrepreneurs in small and medium enterprises in laos and cambodia. *GMSARN International Journal*, *2*(4), 147–156.

Starr, M. A. (2005, September). Does money matter in the CIS? Effects of monetary policy on output and prices. *Journal of Comparative Economics*, *33*(3), 441-461. Retrieved from https://ideas.repec.org/a/eee/jcecon/v33y2005i3p441-461.html

Stevens, N. M. (1905). Studies in spermatogenesis. Retrieved from http://embryo.asu.edu/handle/10776/7511

Stiglitz, J. E., & Weiss, A. (1981, June). Credit Rationing in Markets with Imperfect Information. *American Economic Review*, *71*(3), 393-410. Retrieved from https://ideas.repec.org/a/aea/aecrev/v71y1981i3p393-410.html

Storm, B., Stone, S., & Benjamin, A. (2017, 07). Using the internet to access information inflates future use of the internet to access other information. *Memory*, *25*, 717-723. doi: 10.1080/09658211.2016.1210171

Strain, M. R. (2019, February). *The Link Between Wages and Productivity Is Strong* (John G. Searle Scholar and Director of Economic Policy Studies). American Enterprise Institute (AEI).

Sutton, S. W. (1903). On the morphology of the chromosome group in brachystola magna. *Biol Bull*, 24-39.

Taaffe, J., Morrill, R. L., & Gould, P. R. (1963). Transport expansion in underdeveloped countries: A comparative analysis. *Geographical Review*, *53*, 503–52910.

Takaku, R. (2018). First daughter effects in Japan. *Journal of the Japanese and International Economies*, *50*(C), 48-59. Retrieved from https://ideas.repec.org/a/eee/jjieco/v50y2018icp48-59.html doi: 10.1016/j.jjie.2018.08.00

Temah, C. (2009, 01). What drives hiv/aids epidemic in sub-saharan africa? *Revue d'économie du développement*, *2"*, 41-70. doi: 10.3917/edd.235.0041

Tian, L. (2021, September). *Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil* (CEPR Discussion Papers No. 16590). C.E.P.R. Discussion Papers. Retrieved from https://ideas.repec.org/p/cpr/ceprdp/16590.html

Tollefsen, A., & Buhaug, H. (2015, 02). Insurgency and inaccessibility. *International Studies Review*, *17*, 6-25.

Trabelsi, S., Labelle, R., & Dumontier, P. (2008). Incremental voluntary disclosure on corporate websites, determinants and consequences. *Journal of Contemporary Accounting Economics*, *4*(2), 120-155. Retrieved from https://www.sciencedirect.com/science/article/pii/S1815566910700321 doi: https://doi.org/10.1016/S1815-5669(10)70032-1

Trivers, R., & Willard, D. (1973, 02). Natural selection of parental ability to vary the sex ratio of offspring. *Science (New York, N.Y.)*, *179*, 90-2. doi: 10.1126/science.179.4068.90

UNAIDS. (n.d.).

Vasileiadou, E., & Vliegenthart, R. (2009). Research productivity in the era of internet revisited. *Research Policy*, *38*(8), 1260–1268. doi: 10.1016/j.respol.2009.06.005

Vives, X. (2014). On the possibility of informationally efficient markets. *Journal of the European Economic Association*, *12*(5), 1200–1239. Retrieved from http://www.jstor.org/stable/24539124

Volpe Martincus, C., Carballo, J., & Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. *Journal of Development Economics*, *125*, 21-39. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387816300803 doi: https://doi.org/10.1016/j.jdeveco.2016.10.002

von Broembsen, M., Wood, E., & Herrington, M. (2005, 01). Global entrepreneurship monitor south african report 2005.

Wagner, S., & Cockburn, I. (2010, March). Patents and the survival of Internet-related IPOs. *Research Policy*, *39*(2), 214-228. Retrieved from https://ideas.repec.org/a/eee/respol/v39y2010i2p214-228.html

Waverman, L., Meschi, M., & Fuss, M. (2005, 01). The impact of telecoms on economic growth in developing countries. *London Business School*, *3*.

Wibbels, T., Bull, J., & Crews, D. (1994, 09). Temperature[U+2010]dependent sex determination: A mechanistic approach. *Journal of Experimental Zoology*, *270*, 71 - 78. doi: 10.1002/jez.1402700108

Wilson, E. B. (1905). The chromosomes in relation to the determination of sex in insects. *Science*, *22*(564), 500–502. Retrieved from http://www.jstor.org/stable/1631972

Yamauchi, F. (2016, July). The Effects of Improved Roads on Wages and Employment: Evidence from Rural Labour Markets in Indonesia. *Journal of Development Studies*, *52*(7), 1046-1061. Retrieved from https://ideas.repec.org/a/taf/jdevst/v52y2016i7p1046-1061.html doi: 10.1080/00220388.2015.112

Yamauchi, F., Muto, M., Chowdhury, S., Dewina, R., & Sumaryanto, S. (2011). Are Schooling and Roads Complementary? Evidence from Income Dynamics in Rural Indonesia. *World Development*, *39*(12), 2232-2244. Retrieved from https://ideas.repec.org/a/eee/wdevel/v39y2011i12p2232-2244.html doi: 10.1016/j.worlddev.2011.0

Zhang, X., & Zhang, L. (2015). How does the internet affect the financial market? an equilibrium model of internet-facilitated feedback trading. *MIS Q.*, *39*, 17-38.

Zhu, X. (2012). Understanding china's growth: Past, present, and future. *Journal of Economic Perspectives*, *26*(4), 103–24.