Internet usage and migration decisions: Evidence from Nigerian micro data*

Maurizio Strazzeri†

University of Bern
Department of Economics

January 31, 2021

Abstract

This paper investigates the role of Internet usage in the migration decision using micro-level data from Nigeria. Internet usage reduces migration costs such as search and information costs or psychological costs, which suggests that having access to the Internet increases the probability to migrate. My empirical analysis exploits variation in Internet usage induced by the arrival of submarine Internet cables in Western Africa. Results indicate a large positive effect of Internet usage on migration. The effect is particularly strong for migration out of Africa and is larger for individuals from the lower part of the wealth distribution.

*I thank Luna Bellani, Enzo Brox, Sebastian Findeisen, Tommy Krieger, Stephan Maurer, Guido Schwerdt, and seminar audiences at the University of Konstanz and the ESPE conference 2018 (Antwerp), 2019 (Bath), and the EEA-ESEM conference 2019 (Manchester) for helpful comments and suggestions.

†Email: maurizio.strazzeri@vwi.unibe.ch

1 Introduction

The recent increase in the number of migrants from low and middle income countries to richer countries has led to a renewed interest in the determinants of migration among policy makers and economists alike. It has been noted that advances in modern information and communication technologies might partly explain these migration flows as they reduce the cost of information diffusion and provide potential migrants with information about migration opportunities that were previously unknown (Czaika and de Haas, 2014; Ortega and Peri, 2015).[1] Indeed, the large-scale adoption of the Internet around the world had tremendous impact on the amount, quality, and variety of information that individuals can obtain at relatively low cost. It also facilitated communication with peers over long distances by providing access to modern communication technologies such as emails, social media platforms, or instant messaging technologies. This might have important consequences for complex forward-looking migration decisions that require extensive information-gathering activities and are largely affected by communication via migration networks (Carrington et al., 1996; Munshi, 2003; McKenzie and Rapoport, 2007; Hanson and McIntosh, 2010). Qualitative studies provide evidence that the Internet is used by migrants to inform themselves about, e.g., potential destination countries, migration routes, and immigration regulations, or to maintain connections with peers in the origin country (e.g., Dekker and Engbersen, 2014; Zijlstra and Liempt, 2017), and it has been shown that Internet behavior in origin countries can be used to predict future migration flows (Böhme et al., 2020). However, less is known about whether the Internet is only used as a substitute for other means to implement migration decisions or whether Internet usage directly affects migration decision, i.e., Internet usage makes international migration more likely.

In this paper, I provide the first systematic analysis of the effect of Internet usage on migration decisions based on micro data from Nigeria. The analysis focuses on the years 2010 to

[1]Other potential explanations that have been discussed in literature are economic development and rising income levels in developing countries (Clemens, 2014), climate change which disproportionally affects low income countries (Beine and Parsons, 2015; Cattaneo and Peri, 2016; Missirian and Schlenker, 2017), or demographic imbalances due to a particularly young population in many developing countries (Ortega and Peri, 2015).

2016, where - at the beginning - Internet usage was extremely rare but became more common over time, which provides valuable variation in first-time Internet usage. However, any study of Internet usage and migration is confronted with considerable selection problems as various socio-economic background characteristics that influence the migration decision - such as age, education or income level - affect simultaneously the likelihood of Internet usage. To overcome such endogeneity concerns, I follow Hjort and Poulsen (2019) and exploit time and cross-sectional variation in Internet speed generated by the arrival of large submarine Internet cables from Europe that presumably facilitated Internet usage for individuals living in Nigeria. These submarine cables were connected to the terrestrial cable network of Nigeria between the years 2010 and 2012 and brought much faster Internet and higher Internet traffic capacities to locations close to the terrestrial cable network but not to others. My preferred empirical strategy exploits the arrival of the submarine Internet cables by comparing individuals located close to the terrestrial cable network before and after the arrival of submarine cables with individuals who are located further away during the same time period.

The empirical analysis is based on the comprehensive Nigerian geo-coded General Household Survey (GHS) panel which provides extensive information about individuals' Internet usage behavior. Additionally, the time dimension of the Nigerian GHS panel allows to reconstruct individuals' migration decisions as remaining household members state the whereabouts of former survey participants that moved out of the household. To exploit the variation generated by the arrival of submarine Internet cables, I match the linked information about migration decisions and previous Internet usage with detailed maps of the terrestrial cable network in Nigeria prior to the arrival of the submarine Internet cables.

I start my empirical analysis with various reduced form difference-in-difference estimates using binary or continuous measures of the distance to the terrestrial cable network. I show that individuals located in communities close to the terrestrial cable network respond with a larger increase in migration rates than those individuals located in more remote locations after the arrival of the submarine Internet cables. The change in migration rates for the period before and after the arrival of fast Internet is 0.14 pp lower for individuals located in communities twice as far as the comparison group (approx. -17.7 % relative to mean migration rates in locations close to the cable network). Estimation results using a binary treatment variable suggest that the effect is particularly large for individuals located in a 5 kilometres (km) radius around the terrestrial cable network, which is in line with the hypothesis that individuals are particularly affected if they live close to the terrestrial cable network. Moreover, the absolute effect on migration is twice as large for younger individuals in the age bracket of 20 to 35. As younger individuals are more likely to respond to faster Internet at the extensive (Internet take-up) and intensive margin (more frequent Internet usage), this results suggests that the association between the availability of fast Internet and migration rates is driven by changes in Internet behavior.

To further explore this channel, I show that the arrival of faster Internet is indeed associated with an increase in Internet usage at the extensive and intensive margin. With respect to the extensive margin, being located within a 5 km radius around the terrestrial cable network is associated with an increase in the probability to use the Internet by around 5 pp for the entire sample and 10 pp for younger individuals after the arrival of the submarine Internet cables in comparison to individuals located in more remote locations. Under the assumption that the exclusion restriction is satisfied - which I extensively test in the paper - these results indicate a large effect of Internet usage on migration decisions. Two-stage least squares estimates suggests that Internet usage increases the probability to migrate by 10 pp.

The positive effect of Internet usage on migration is consistent with a standard model of migration where individuals compare benefits and costs associated with moving (Sjaastad, 1962; Borjas, 1987), and Internet usage lowers the cost of migrating. Internet usage might reduce migration costs by lowering search and information costs or psychological costs that "incur due to the reluctance of individuals to leave familiar surroundings, family, and friends" (Sjaastad, 1962). It might also be the case that the exposure to foreign media that is more prevalent on the Internet changes individuals' preferences for one country in comparison to the origin country such as preferences for a particular climate or lifestyle. All these channels suggest that the effect of Internet usage on migration decisions varies across observable characteristics of destination countries and potential migrants. My empirical set-up allows to assess such heterogeneous effects and I highlight two important consequences of higher migration rates that are caused by increased Internet usage.

First, I show that the effect of Internet usage on migration decisions is more important for migration out of Africa in comparison to migration within Africa, which suggests that the reduction in migration costs induced by Internet usage is larger for extra-continental destination countries. This is in line with the hypotheses that Internet usage lowers search and information costs and changes the set of information available to potential migrants, which might be particularly important for migration out of Africa, or that the exposure to foreign media transmitted via the Internet affect individuals' preferences. Second, I show that the effect of Internet usage on migration differs by individuals' wealth, and individuals from the lower part of the wealth distribution respond with a higher increase in migration rates due to Internet usage. Assuming that individuals' wealth is associated with skills (Angelucci, 2015), this result highlights the self-selection and resulting skill distribution of immigrants caused by the spread of the Internet. A potential explanation for the differential change in migration costs due to Internet usage might be that the migration related information provided by the Internet is more valuable for individuals at the lower part of the skill distribution. Additionally, these individuals are more often confronted with financial constraints that prevent them from migrating (Angelucci, 2015; Bazzi, 2017), which suggests that reduction in migration cost due to Internet usage relaxes these previous constraints and low-skilled migration becomes more likely.

In the final part of this paper, I complement my main findings with an investigation of possible feedback effects of the increased migration rates due to higher Internet usage. The impact of international migration on the development in sending countries has received increasing attention in the literature (Rapoport and Docquier, 2006; Yang, 2011; Docquier and Rapoport, 2012). In particular, many scholars have argued that remittances sent by migrants are an important element for the well-being of the household members left behind. My empirical investigation supports this hypothesis at least for some measures of economic development. I show that households in locations that saw a relative increase in migration rates due to the arrival of the submarine Internet cable are also more likely to report to have received remittances in the following period. However, while this increase in remittances is accompanied by increased investment in secondary education of those left behind, I do not find any positive effects on household wealth. While I cannot rule out direct effects of the arrival of fast Internet on remittances, e.g., by providing a better infrastructure for money transfers, I tentatively interpret these findings as positive feedback effects of the Internet-induced increase in migration rates on the economic development in Nigeria.[2]

In sum, this paper provides novel insights into the effect of Internet usage on migration decisions, and highlights potential consequences of the spread of the Internet on the direction of international migration flows, self-selection of migrants, and the skill distribution of immigrants in receiving countries. Therefore, this paper contributes to the vast empirical literature on the determinants of migration (e.g., Hatton, 2005; Mayda, 2010; Ortega and Peri, 2013) and, in particular, to studies that investigate migration costs and the skill distribution of migrants. These studies exploit changes in migration costs due to, e.g., pre-existing migration networks (Munshi, 2003; McKenzie and Rapoport, 2007), immigration and border controls (Angelucci, 2012; Allen et al., 2018), or cultural and linguistic differences between destination and origin countries (Belot and Ederveen, 2012; Adsera and Pytlikova, 2015). In line with the application of this paper, Feigenberg (2020) highlights differential changes in migration costs for different types of potential migrants due to the United States-Mexico border fence construction that disproportionally reduces migration from low skilled migrants. This paper also relates to the literature that highlights wealth and income constraints which might deter low skilled migration (Dustmann and Okatenko, 2014; Bazzi, 2017; Cai, 2020). For instance, Angelucci (2015) shows that poor households in Mexico that experience an exogenous increase in income are more likely to migrate which worsens skills among Mexican migrants in the United States. This paper also contributes to the literature on the effect of media exposure on various socio-

[2]In a recent study, Lee et al. (2020) show that Internet based mobile technology significantly increases urban-to-rural remittances, which suggests that direct effects of the arrival of fast Internet on remittances might be a concern here.

economic outcomes.[3] In a related study, Farré and Fasani (2013) link TV usage to internal migration decisions in Indonesia and find, contrary to this study, that TV usage reduces the likelihood of migrating longer distances within Indonesia. The authors explain their finding by arguing that Indonesians, on average, over-estimate the returns to internal migration when they only have limited access to television. On the other hand, Braga (2007) finds that Albanians who were exposed to Italian television are more likely to migrate internationally. She argues that exposure to Italian television increased the availability of information about various lifestyles of societies in the Western world which might have affected the aspirations of migrants. Two other studies link advances in communication technologies to migration decisions. Lu et al. (2016) show that the installation of landline phones in rural China intensified internal migration by providing better access to information about job opportunities due to stronger migration networks and reducing the psychological costs of migrating. Aker et al. (2011) also stress the importance of information provision via mobile phones in migration decisions and show that adult education programs in Niger in which participants learned to use mobile phones led to an significant increase in seasonal migration.

The rest of the paper is organized as follows. In Section 2, I provide background information on international migration in Nigeria (Section 2.1) as well as the Nigerian Internet infrastructure and the arrival of the large submarine Internet cables starting in the year 2010 (Section 2.2). In Section 3, I present my data set. In Section 4, I introduce the main identification strategy. The baseline results and various robustness test are in Section 5, and in Section 6, I provide a discussion of the heterogeneous effects of Internet usage on migration. In Section 7, I discuss potential feedback effects of increased migration due to Internet usage, and Section 8 concludes.

[3] Media exposure has been linked to, e.g., educational outcomes (Gentzkow and Shapiro, 2008), crime rates (Dahl and DellaVigna, 2009), and fertility decisions (La Ferrara et al., 2012). For an extensive survey on media exposure, see DellaVigna and La Ferrara (2015).

2 Background

2.1 International migration in Nigeria

Nigeria is a multinational state in Western Africa that was formed under British colonial rule in the beginning of the nineteenth century and is today inhabited by hundreds of different ethnic groups who speak more than 500 different languages. Commonly referred to as the "Giant of Africa," it is is the most populous country in Africa with a population of more than 200 million people, which represents around one-sixth of the continent's population. It has experienced a massive population growth, having had only 45 million inhabitants in the 1960s, and due to its young population - more than 50 % of Nigerians are below the age of 18 - and high fertility rates it is projected to grow even further. The United Nations Department of Economic and Social Affairs estimates that by 2050 Nigeria's population will surpass that of the United States, and will increase to more than 700 million people by the end of the 21th century (UN DESA, Population Division, 2019). Nigeria has an abundance of natural resources and the biggest oil and natural gas resources on the African continent. It's economy and public finances depend heavily upon oil revenues. As a result, economic growth has been exceptionally volatile, especially during the oil price crises in the 1970s. Since Nigeria's independence in 1960 to the first general election in 1999, the Nigerian government was almost exclusively under the rule of military dictators. It has experienced a number of national conflicts, many of which have arisen from ethnic or religious conflicts. While a number of social and economic indicators such as life expectancy at birth and years of schooling have improved over the last two decades (UNDP, 2019), many Nigerians still suffer from extreme poverty as more than 50 % of people are below the international poverty line set by the Worldbank (WorldPovertyClock, 2020).

Due to its massive size and it's economic and demographic peculiarities, Nigeria has played an important role in African migration, and will likely do so in the years to come.[4] Over the past decades, from its independence to today, Nigeria experienced a "reverse migration transi-

[4]This paragraph draws heavily on De Haas (2007) who provides an excellent short overview on the history of international migration in Nigeria.

tion" as it transformed from a major immigration country among Western African countries to a net emigration country (Black et al., 2004). In particular during the 1970s when Nigeria saw an increase in its oil revenues which lead to higher incomes among middle class households, it attracted a large number of labor migrants from neighbouring countries. During this time, most Nigerians who left the country did so only for study or business reasons, and primarily went to the United Kingdom or the United States. Although this type of high skilled migration to Anglo-Saxon countries continued for the following decades, cross-border migration in Nigeria became more diverse, intense, and permanent during the 1980s when decreases in oil prices lead to a long economic downturn and political repression and violence became more widespread (Hernández-Coss and Egwuagu Bun, 2007). High demand for less skilled workers in Western European countries such as Spain, Italy, Germany and France, the Gulf states, and more wealthy economies in Africa (e.g., South Africa, Botswana, Gabon) attracted more and more workers from Nigeria (Black et al., 2004). Increasing immigration restrictions among European countries implemented over time did not lower these migration flows but instead amplified the amount of irregular migration. While until the 1990s the majority of migrants used air links, the means of travelling changed and a considerable amount of migration from Nigeria to Europe is by now trans-Saharan (De Haas, 2006).

Today, a significant number of Nigerians who reach European terrain apply for asylum in Italy, Germany, France or the UK. Today, in absolute numbers, Nigerians are the largest group of asylum seekers from Western African countries in Europe, and the number of Nigerian asylum seekers has been growing over the years, even though rejection rates among Nigerian asylum seekers are relatively high.[5] The inflow of migrants from Nigeria to Europe might not stop any time soon as the willingness to leave the country seems still to be considerably high. In a survey conducted by the PEW Research Center, more than two-thirds of the respondent

[5]Between 2010 and 2015, around 86 thousand Nigerians applied for asylum in the European Union which represents around 33 % among all applicants from the 15 Economic Community of Western African States (ECOWAS) countries. The absolute change in asylum applications from Nigerians between 2010 and 2015 is the highest among all other ECOWAS countries (23 thousand), and the relative change to the year 2010 (357 %) is only larger for Gambians (995 %) and Senegalese (657 %). Own calculations based on data from the United Nations High Commissioner for Refugees. Around 6.6 % of asylum seekers from Nigeria who received the decision about their asylum application between January and October 2019 received some form of international protection in Germany (BAMF, 2019).

stated that they would leave Nigeria if they had the means and opportunity to do so, and almost 40 % say they plan to leave Nigeria within the next five years (Connor, 2018). Similar results can be obtained using the last wave of the Afrobarometer conducted in the year 2017 where around 25 % of the Nigerian respondents reported to have considered to move to another country, of whom 48 % are planning to move within the next two years (36 %) or are currently making preparations to move (12 %), and around 80 % would most likely go to a country outside of the African continent.

2.2 Nigerian Internet infrastructure and the arrival of fast Internet

The recent increase in the number of Internet users in Sub-Saharan African countries provides an interesting setting for investigating the impact of the exposure to the Internet on migration behavior. From being basically non-existent in the year 2000, the share of Internet users began slowly to grow over the following years, with approximately 7 % of individuals using the Internet in the year 2010 to around 25 % in the year 2017.[6] As the largest country in Sub-Saharan Africa, much of this growth in Internet usage was driven by Internet users in Nigeria, where latest numbers suggest that already around 125 million people use the Internet.[7]

Internet usage in Nigeria - as in other Sub-Saharan Africa countries - depends heavily on submarine Internet cables, which are globally responsible for about 99 % of international communication traffic (Brake, 2019).[8] Submarine Internet cable are fiber glass cables that are laid on the sea bed and carry telecommunication signals across oceans. They provide the necessary link between end users in Nigeria and Internet content provider that are not hosted on the African continent or, in most cases, even in Nigeria.[9] Submarine Internet cables are particularly important for African Internet users as only the minority of Internet content is hosted locally, and, in many cases, even local content is hosted overseas due to lower costs (Kende

[6]In comparison, the share of Internet users in the European Union (United States) grow from 20 % (43) in the year 2000 to 68 % (70) in the year 2010 and 82 % (87) in the year 2017.

[7]Source: https://www.internetworldstats.com/africa.htm.

[8]While possible, only a minority of international communication is carried out via satellites, which is slower and more expensive. An overview on the economics of Internet infrastructure gives Greenstein (2020).

[9]There is almost no connection between landline networks in Africa, which implies that international Internet traffic in Africa needs to travel overseas (Hjort and Poulsen, 2019).

and Rose, 2015).[10] The first submarine Internet cable that provided Nigeria with a connection to the Internet is the *SAT-3* cable, which begins in Sesimbra, Portugal, and was connected to the port of Lagos in the year 2001 and finally started to operate in 2002 (Stanley et al., 2018). With a capacity of 340 Gbit/s, it served as the only source of Internet connectivity at this time and damages to the *SAT-3* cable were responsible for tremendous internet blackouts. For instance, when in July 2009 the *SAT-3* cable was accidentally cut, Nigeria's Telecommunication operators were forced to use satellite links to maintain Internet connectivity which reduced available bandwidth in Nigeria by 70 %, causing problems to various sectors in the economy as well as individual end users.[11]

Starting in the year 2010, four new submarine Internet cables were connected from Europe to land-based stations in Western Africa, bringing more reliable Internet connectivity, larger Internet traffic capacities, and lower bandwidth prices to end users in Nigeria. The first cable called *Main One* was connected in July 2010 to a land-based station in Lagos, Nigeria, with a capacity of 1.28 Tbit/s at time of installation that can be extended to up to 4.96 Tbis/s.[12] The availability of larger Internet traffic capacities due to the arrival of the new submarine Internet cables increased Internet speed for end user in Africa by around 35 % (Hjort and Poulsen, 2019). Additionally, the *Main One* cable had a significant impact on bandwidth prices, contributing to "an immediate drop of 50 % on the price of bandwidth in Nigeria" (Stanley et al., 2018). Lower prices and faster and more reliable Internet connection suggests that Internet usage became more convenient in Nigeria after the arrival of the *Main One* cable, with potential positive effects on Internet usage and Internet usage frequency (Hjort and Poulsen, 2019). However, it is important for the identification strategy of this paper that not all Nigerians immediately benefited from the arrival the new submarine Internet cables. After being plugged in to a land-based station, these submarine Internet cables brought faster Internet traffic capacities only to locations that are connected to the national terrestrial cable network.[13] The connection

[10]Chavula et al. (2014) estimate that, on average, more than 75 % of traffic that originates from African universities are transmitted outside the continent.

[11]BBC News, 30 July 2009. http://news.bbc.co.uk/2/hi/technology/8176014.stm (accessed: 2020-09-09).

[12]https://www.mainone.net/our-network-3/cable-system.

[13]Internet traffic within Africa is transmitted via the telephone cables that were built many decades back (Hjort and Poulsen, 2019). In Section 3, I provide a discussion of the diffusion of the terrestrial cable network in Nigeria.

between the terrestrial cable network and the end user, the so called "last mile" technology, might be *wireline* through fiber or copper cables, or *wireless* using cell towers or satellites. Both the type of the last mile technology and the distance to the terrestrial cable network determine the experienced Internet speed of end user (Greenstein, 2020). As Internet access via mobile phones and cell towers was very rare at that time, it is likely that only locations in close distance to the terrestrial cable network benefited from the arrival of the submarine Internet cables.[14] Technical considerations suggest that fast Internet might be available only within 500 meters distance to the terrestrial cable network for last mile technologies based on copper cable (Hjort and Poulsen, 2019). However, potential spill-over effects to adjacent locations are likely as most Internet users in Nigeria had access to the Internet only via Internet cafes.[15] I provide a discussion of such potential spill-over effects in Section 4.

3 Data and descriptive statistics

The main data source of this paper is the geo-coded Nigerian General Household Survey (GHS) panel which is administered by the National Bureau of Statistics of the Federal Republic of Nigeria.[16] The first wave of the GHS panel is a sample of 5,000 households with more than 27,000 individuals based on 500 enumeration areas (communities) who were interviewed in the years 2010, 2012, and 2016. For the empirical analysis, I use a balanced sample at the household level (4,407 households) and exclude individuals who were below the age of 15 or above the age of 65 in the first wave.

To construct the main explanatory variables, I use information from the *ICT Usage Section* of the GHS panel interview questionnaire. In particular, I use the question: "Do you have access to the Internet?" to construct a binary measure of Internet usage and the follow-up question: "How often do you use the Internet?" to construct an ordinal measure of Internet

[14] 3G services have only been introduced in 2007 in Nigeria and the share of subscribers to 3G or 4G services in 2010 was close to zero (GSMA, 2015).

[15] Around 80 % or participants in the Nigerian General Household Survey panel who reported to have access to the Internet in the year 2010 say their main access to the Internet is through Internet cafes. For studies that highlight the importance of Internet cafes in Sub-Saharan Africa, see, e.g., Mwesige (2004) or Adetoro (2010).

[16] The data is publicly available and can be obtained from the website of the World Bank Microdata Library: http://microdata.worldbank.org.

usage frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily).

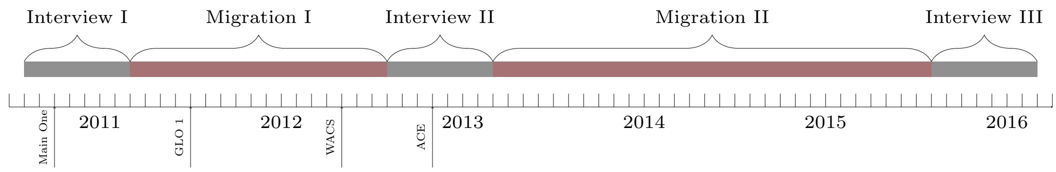
To construct the main outcome variable of this paper, I exploit the panel structure of the data. If a household member moves out of the household between two subsequent waves, the remaining household members report the whereabouts of the individual, i.e., the remaining household members state if an individual left the country or moved within Nigeria. I link the information about the migration decision provided by the remaining household members to the information the respondent provided in the previous wave. The main outcome variable *international migration* is equal to one if the individual moved out of Nigeria and zero otherwise. Note that, as the GHS panel does not provide information about migration behavior after the last (third) wave, the main empirical analysis will be based on the first and second wave of the GHS panel.

Basic background information such as age, gender, education, and other ICT usage such as TV usage and mobile phone usage is also available in the GHS panel. To obtain a measure of wealth, I follow Young (2013) in the context of developing countries, and use information about four housing conditions to obtain an ordinal measure of wealth, i.e., the wealth measure is equal to the number of housing conditions that are fulfilled (household conditions: (i) constructed floor made of other than dirt, sand or dung, (ii) flush toilet, (iii) tapped drinking water, (iv) electricity in house).

I link the data from the GHS panel with detailed maps of the terrestrial cable network prior to the arrival of the submarine Internet cables. I obtain the data for the terrestrial cable network in Nigeria from Hjort and Poulsen (2019) who use this information in a related study.[17] The GHS panel provides information about the location at the community level. I use this information as well as the information from the location of the terrestrial cable network to calculate the shortest distance between the community and the terrestrial cable. Figure A1 in the Appendix shows the terrestrial cable network (red solid lines) on a map of Nigeria as well as the communities included in the final GHS data set. Figure A1 illustrates the sparse diffusion of

[17]Hjort and Poulsen (2019) obtained the data from www.africabandwidthmaps.com and www.afterfibre.net.

Figure 1:
Time line of events: Arrival of submarine Internet cables and survey waves



Sources: Stanley et al. (2018), Hjort and Poulsen (2019), Nigerian GHS panel.

the terrestrial cable network in Nigeria, with many locations - in particular in Western Nigeria - being many hundreds of kilometres away from the terrestrial cable network. However, as shown in Figure A2 in the Appendix, which provides a magnified view on Southern Nigeria, there are also a large number of communities in the GHS panel that are relatively close to the terrestrial cable network.

In Figure 1, I provide an overview of the arrival of the submarine Internet cables in Nigeria as well as the interview periods of the Nigeria GHS panel.[18] As explained above, to analyze the effect of Internet usage on migration, I match information given by remaining household members from the subsequent interview to the socio-economic information from the previous wave. Ideally, for the empirical strategy outlined below, I would like to have at least one period before the arrival of fast Internet in Nigeria and some periods after. Unfortunately, as can be seen in Figure 1, the first submarine cable was connected already during the first interview period. If the boost in Internet speed already lead to increased Internet usage in the first period in locations close to the terrestrial cable network, defining the first wave of the GHS panel as untreated in a difference-in-difference approach would contaminate the estimation results and bias the estimates towards zero. However, it seems unrealistic that the arrival of submarine Internet cables affected Internet usage behavior in such a short time period. Accordingly, and considering that in case of misclassification the difference-in-difference approach provides conservative estimates of the true effect, I define the first wave of the GHS panel as unaffected by the arrival of fast Internet in Nigeria throughout the paper.

[18]Unfortunately, the GHS panel does not provide information about the exact day of interview.

Table 1:
Mean values of selected variables by Internet usage

	No Internet usage	Internet usage
Socio-economic		
Age	32.56	29.13
Female	0.53	0.35
Household member		
Head	0.26	0.23
Spouse	0.34	0.09
Son/Daugther	0.35	0.61
Other	0.05	0.08
Currently enrolled	0.22	0.45
Highest education		
No schooling	0.37	0.02
Some schooling	0.25	0.04
Secondary education	0.35	0.66
University degree	0.02	0.28
Number of wealth items		
0	0.25	0.02
1	0.31	0.06
2	0.28	0.31
3	0.14	0.53
4	0.02	0.09
Other ICT usage		
Television	0.51	0.97
Mobile phone	0.80	0.99
Location		
Urban	0.24	0.64
Distance next road	14.62	7.33
Internet usage frequency		
At least once a month	0.00	0.34
At least once a week	0.00	0.45
Daily	0.00	0.21
Outcome		
International Migration (in %)	0.18	1.23
Observations		
Total	20,328	1,298
Share in 2012	0.50	0.60

Note: Mean values of covariates by Internet usage. Number of wealth items based on housing conditions: Constructed floor (made of other than dirt, sand or dung), flush toilet, tapped drinking water, electricity in house.

In Table 1, I provide an overview of the selected covariates for the pooled data set by the binary measure of Internet usage. Table 1 illustrates - as expected - stark differences between Internet user and non-Internet users, while still showing at least some overlap across treatment status. On average, individuals who have access to the Internet are younger and less likely to be female, are more likely to be enrolled in school, and have a higher education at interview date. Unsurprisingly, usage of other ICT such as TV and mobile phones are more common among

Internet users, and members of this group live in wealthier households which are located more often in urban areas. Contrary to the experience in more developed countries, having access to the Internet does not mean daily usage in Nigeria in the years 2010 and 2012. Only one-fifth of the individuals of this group use the Internet daily, while around one-third access the Internet less than a week.

The lower part of Table 1 also illustrates stark differences in the propensity to migrate between Internet user and non-Internet user. Table 1 suggests that international migration is almost seven times as likely for Internet users than for non-Internet users. As shown in Table A1 in the Appendix, which reports OLS estimates of the binary variable measuring migration on Internet usage, this difference cannot be explained by the factors listed in Table 1 or unobserved location-specific heterogeneity across states or counties. However, the OLS estimates reported in Table A1 cannot be considered as causal effects of Internet usage on migration. The large differences in observable characteristics between both groups suggests that there are also considerable differences in unobservable variables between both groups that might be related to the decision to migrate, which, in turn, would bias these estimates.

4 Identification

4.1 Effect of the arrival of fast Internet on migration

The empirical approach in this paper exploits time and cross-sectional variation generated by the arrival of the first submarine Internet cable from Europe. Individuals living close to terrestrial cable network experienced significantly higher Internet speed after the arrival of the submarine Internet cable in comparison to the years before. On the contrary, individuals living in more remote locations were not affected by the arrival of the submarine Internet cable. If the availability of fast Internet affected Internet take-up and Internet usage frequency, which in turn influenced migration decisions, I expect that changes in migration rates before and after the arrival of the first submarine Internet cable are larger in locations close to the terrestrial cable network. More formally, to exploit the variation generated by the arrival of fast Internet, I estimate the following fixed effect specification:

$$Migration_{i,c(i),t+k(t)} = \mu_{c(i)} + \beta_0 \mathbb{1}\left[t = 12\right] + \beta_1 \mathbb{1}\left[t = 12\right] * Distance_{c(i)} + v_{i,c(i),t}, \quad (1)$$

where $Migration_{i,c(i),t+k(t)}$ is a binary variable indicating whether individual i from community $c(i)$ moved to another country before time $t + k(t)$, where $k(t)$ represents the number of years between two successive waves of the Nigerian GHS panel. $Distance_{c(i)}$ is a measure of the distance of individual i's community $c(i)$ to the terrestrial cable network. $\mathbb{1}\left[\cdot\right]$ is an indicator function equal to one if the condition in the brackets is fulfilled, i.e., $\mathbb{1}\left[t = 12\right]$ is equal to one in year 2012 and zero otherwise.[19] $\mu_{c(i)}$ represents a set of community fixed effects and $v_{i,c(i),t}$ is an error term which captures all effects that influence $Migration_{i,c(i),t+k(t)}$ that are not caused by other factors included in Equation (1).

I estimate Equation (1) based on data for the years 2010 and 2012.[20] The inclusion of community fixed effects allows for any systematic time-invariant variation in migration behavior across locations due to, e.g., geographic characteristics, pre-existing migration networks, or income levels. On the other hand, the inclusion of a year dummy allows for systematic variation in migration behavior after the first and the second wave that affect all locations such as varying time windows for migration between the first and second wave and the second and third wave of the Nigerian GHS panel or time-specific economic shocks in origin or destination countries influencing migration behavior of all locations in the sample. Consequently, the coefficient of interest in Equation (1), β_1, can be interpreted as a difference-in-difference estimator. This means, β_1 measures the difference in migration rates between the time windows 2010-12 and 2012-15 between (i) communities that marginally differ in $Distance_{c(i)}$ for a continuous measure of $Distance_{c(i)}$ or (ii) connected ($Distance_{c(i)} = 1$) and unconnected locations ($Distance_{c(i)} = 0$) for a binary measure of $Distance_{i,c(i),t}$. While a binary measure facilitates the interpretation of β_1, it is challenging to determine the radius around the terrestrial cable network that defines treated and untreated locations. Based on technical con-

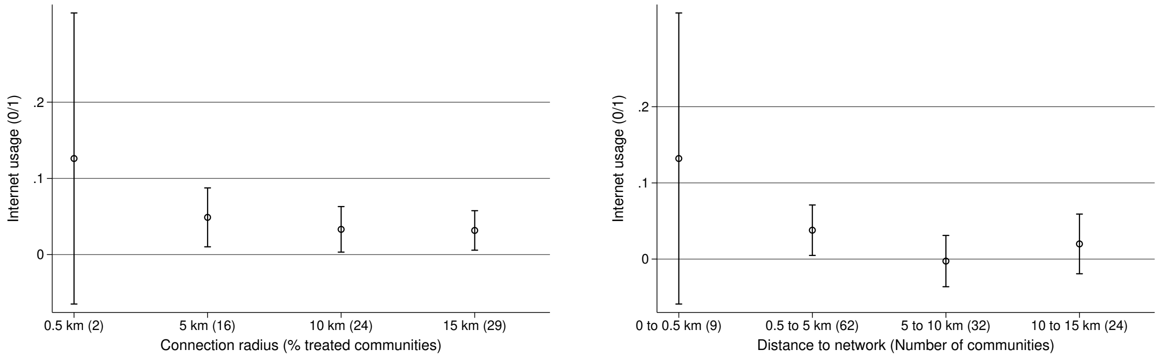
[19] All equations in this paper use the years since 2000 to refer to a given year.

[20] As explained in more detail above, information about the migration decision is always obtained from the following wave. See also Figure 1 and corresponding discussion in Section 3.

siderations, Hjort and Poulsen (2019) treat locations within a 0.5 km radius as connected to the terrestrial cable network. The authors argue that the 500 meters radius around the terrestrial cable network is a good proxy for the availability of fast Internet in a location as the transmission rate of "last-mile" technologies based on copper cables - which are prevalent in most African countries - become significantly lower beyond this threshold. However, the distance-connectivity relationship might differ tremendously for "last-mile" transmission via microwaves. Additionally, assuming that the availability of fast Internet is indeed restricted to buildings within a 0.5 km radius, significant spill-over effects to individuals living in adjacent locations might be likely as most Nigerian Internet users do not access the Internet via home based Internet connection. This is illustrated in Figure 2, which shows coefficient plots for regressions of Equation (1) when using individual i's Internet usage as dependent variable as well as a binary measure of distance based on various definitions of connected areas. The left plot of Figure 2 shows coefficient estimates for four separate regressions, where the binary variable takes on the value one if individual i is located either within a 0.5 km, 5 km, 10 km or 15 km radius around the terrestrial cable network. Indeed, the increase in Internet usage between connected and unconnected areas caused by the arrival of fast Internet is largest if I employ the most narrow definition for connected areas - even though insignificant possibly due to small sample size, and the effect becomes smaller if the definition becomes wider. To assess the spill-over effects, the right plot of Figure 2 shows coefficients of a similar regression but using a set of binary variables that exclusively define individuals' location (baseline: distance to terrestrial cable network larger than 15 km). This plot shows that I obtain a significant positive effect of the arrival of fast Internet also for individuals living within a corridor of 0.5 to 5 km around the terrestrial cable network, which supports the idea of positive spill-over effects to individuals living in adjacent location. However, this effect already disappears for individuals living in more remote locations.[21] To account for the unknown drop in transmission rate and potential spill-over effects, I use the logarithm of the distance to terrestrial cable

[21]The results shown in Figure 2 are qualitatively similar when using Internet usage frequency as dependent variable or restricting the sample to younger individuals, for which I expect even larger effects. See Figure A3 in the Appendix.

Figure 2:
Internet usage and distance to terrestrial cable network, 2010-12 change



Note: Plot on the left shows coefficient estimates for four separate regressions of Internet usage on an interaction term of a binary variable indicating if individual i is located in a community within the connection radius shown on the x-axis to the terrestrial cable network and an indicator variable for the year 2012. Plot on the right shows coefficient estimates for a regression of Internet usage on a set of binary variables indicating if individual i is located within a bin shown on the x-axis (baseline: Distance to terrestrial cable network larger than 15 km). All estimates include a year dummy for the year 2012 as well as community fixed effects. Number of observations: 21,626. 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster). Figure A3 in the Appendix shows equivalent plots for Internet usage frequency and a sample of younger individuals.

network for $Distance_c$. In Section 5.1, I show that this specification provides a reasonably good fit to the change in Internet usage and Internet usage frequency due to the arrival of fast Internet. Additionally, I report results for a binary measure of distance defining those locations as connected that are within a 5 km radium around the terrestrial cable network.

The arrival of fast Internet affected those individuals stronger who live closer to terrestrial cable network. Hence, if the arrival of faster Internet has a positive effect on Internet usage and subsequently on migration, I expect β_1 to be negative when using the continuous measure of distance and positive in the case of the binary measure. Moreover, I expect the effect to be significantly larger for younger individuals as they are more likely to change their Internet behavior as a response to faster Internet. For this reason, I also report results for the estimation of Equation (1) on a sub-sample of individuals between the age of 20 and 35.

If the common trend assumption is satisfied, β_1 measures the effect of the exposure to fast Internet on migration decisions for individuals located in connected areas. The common trend assumption states that migration rates in communities would have evolved similarly across locations between the time windows 2010-12 and 2012-15 if the submarine cables had not been connected to Nigeria (see, e.g., Abadie, 2005). In Section 5.1, I provide an extensive discussion of possible violations of this parallel trend assumption, as well as of the robustness of my estimation strategy to various specification. I show that my estimation results are robust to (i) using a binary measure of treatment defining all locations as connected if they are within a 5 km radius around the terrestrial cable network, (ii) the exclusion of locations that are close or far away from treated locations, and (iii) the inclusion of various placebo treatments based on interactions of $\mathbb{1}[t = 12]$ with various pre-treatment level-differences between treated and untreated locations. Additionally, I assess pre-treatment trends of the outcome variable between connected and unconnected locations.

4.2 Effect of Internet usage on migration

As the Nigerian GHS panel provides information about Internet usage, I also show instrumental variable estimates of the effect of Internet usage on the decision to migrate, exploiting the plausible exogenous variation in Internet usage caused by the arrival of fast Internet. To ease the motivation of this approach, I assume to have a binary measure of distance to the terrestrial cable network that assigns individuals to locations that either receive fast Internet after the arrival of submarine Internet cables (connected locations) or not (unconnected locations). The share of Internet users increased due to the arrival of submarine Internet cables more in connected locations than in other unconnected locations. However, the arrival of fast Internet did not incentivize all individuals located in connected areas to use the Internet, and some individuals in unconnected locations might have started to use the Internet for other reasons. Assuming that the increase in Internet usage is larger in connected locations than in unconnected locations, such a fuzzy design still allows to estimate a local average treatment effect (LATE) for a subpopulation of my sample. Following De Chaisemartin and d'Haultfoeuille (2018), this subpopulation consists of *switchers* in connected locations, i.e., individuals located close to the terrestrial cable network (connected locations) that started to use the Internet only after the arrival of the submarine Internet cables. Under additional assumptions stated below, I can obtain the LATE for switchers in connected locations with the following two stage least squares (2SLS) estimation:

$$Internet_{i,c(i),t} = \mu^F_{c(i)} + \beta^F_0 \mathbb{1}\left[t = 12\right] + \beta^F_1 \mathbb{1}\left[t = 12\right] * Distance_{c(i)} + e_{i,c(i),t}, \tag{2}$$

$$Migration_{i,c(i),t+k(t)} = \mu^S_{c(i)} + \beta^S_0 \mathbb{1}\left[t = 12\right] + \beta^S_1 \hat{Internet}_{i,c(i),t} + \epsilon_{i,c(i),t}, \tag{3}$$

where $Internet_{i,c(i),t}$ is a either a binary variable indicating Internet access of individual i in community $c(i)$ at time t or an ordinal measure of Internet usage frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). $\hat{Internet}_{i,c(i),t}$ is the predicted value of Internet usage or Internet usage frequency based on the estimated parameters of Equation (2). $e_{i,c(i),t}$ and $\epsilon_{i,c(i),t}$ are error terms intended to capture effects on the outcome variables that are not caused by factors included in Equation (2) or (3), respectively. All other variables are defined as before.

Equation (2) is the first stage equation and provides estimates of the impact of the arrival of fast Internet on Internet usage or Internet usage frequency. As in the reduced form equation explained above, Equation (2) includes a set of community fixed effects ($\mu^F_{c(i)}$) that allow for systematic time-invariant variation in $Internet_{i,c(i),t}$ across locations as well as a year dummy ($\mathbb{1}\left[t = 12\right]$) to control for an average time trend in Internet usage affecting all locations. The coefficient of interest in this 2SLS procedure is β^S_1 which measures the effect of Internet usage on the migration decision. Estimating β^S_1 by a 2SLS procedure as outlined in Equations (2) and (3) is equivalent to running two separate regressions of Equations (1) and (2) and dividing the estimated parameter β_1 from the reduced form equation by β^F_1 from the first stage equation.

De Chaisemartin and d'Haultfoeuille (2018) outline the assumptions when such a Wald difference-in-difference estimator represents an estimate of the LATE for switchers in connected locations in a fuzzy design. First and most importantly, the common trend assumption needs to be fulfilled. In this setting, the common trend assumption states that the migration rates would have evolved similarly if the share of Internet users had not expanded differently across locations. This is a more restrictive assumption than the one stated above as it assumes that there is no growth (or decline) in other observable or unobservable factors - except of Internet usage - between the two time periods that influence migration rates while the previous common trend assumption allows for changes in factors that influence migration rates if they are caused by the arrival of fast Internet.[22] On the other hand, this assumption is less restrictive than requiring the common trend assumption from the reduced form model also on the treatment variable Internet usage. The share of Internet users are allowed to evolve differently across locations even in the absence of the arrival of fast Internet if the factors determining the different time trends - such as, e.g., previous share of Internet users - do not affect migration rates.

De Chaisemartin and d'Haultfoeuille (2018) further show that the identification of the LATE for switchers in connected locations requires that (i) the treatment effect of Internet usage on migration is stable over time for all individuals and (ii) the treatment effect of Internet usage on migration is homogeneous for switchers in connected and unconnected locations. In particular, the first assumption seems to be problematic as the information provided by the Internet changes over time which might also lead to heterogeneity of the effect of Internet usage on migration. However, the time periods considered here are rather small, suggesting that this might not be a major concern.

I provide a discussion of possible violations of the common trend assumption as well as the robustness of my estimation strategy to various specifications in Section 5.2. In addition to the sensitivity checks mentioned in Section 4.1, I provide a thorough discussion of possible violations of the exclusion restriction of the instrument used in the empirical analysis.

5 Results

5.1 Reduced form estimates

In this section, I provide a discussion of the results for the reduced form relationship between migration decisions and the arrival of fast Internet in Nigeria, and start with a visual inspection

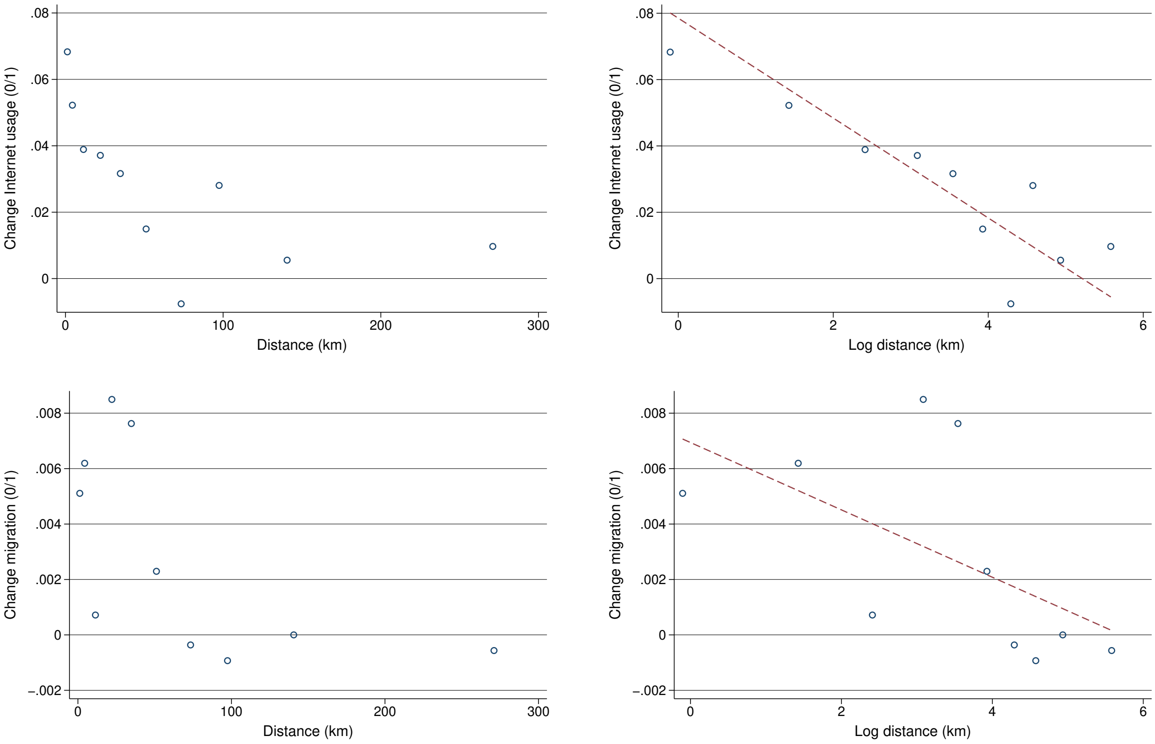
[22]Hence, the common trend assumption stated in this subsection involves an exclusion restriction on the excluded instrument, $\mathbb{1}\left[t=12\right] * Distance_{c(i)}$, the interaction of the year dummy with the distance to the terrestrial cable network measure.

of the correlation between migration rates, Internet usage and the distance to the terrestrial cable network. As the distance to the terrestrial cable network is measured at the community level, estimations of Equation (1) (and (2)) based on individual times year level data are very similar to regressions based on community times year level data where the outcome variable is the share of individuals who migrate (or use the Internet) in a community and year.[23] Additionally, having only two time periods for the main part of the analysis, the fixed effect specification gives equivalent results to a cross-sectional regression based on first differences. Therefore, plotting changes of mean migration rates or mean Internet usage in a community over the two time periods against the distance to the terrestrial cable network provides a clear visual representation of the reduced form relationships discussed above. Figure 3 shows binned scatter plots (10 equally sized bins) of the change in the average Internet usage (upper two plots) and migration rates (lower two plots) in a community between the treated year 2012 (after the arrival of the submarine cables) and the untreated year 2010 (before the arrival of the submarine cables) and the distance to the terrestrial cable network. The distance to the terrestrial cable network is measured in kilometres (km) in the plots on the left and in log in the plots on the right.

The arrival of fast Internet by the connection of the submarine Internet cables was transmitted by the terrestrial cable network and brought faster speed and traffic capacities to locations close to the pre-existing terrestrial cable network. In line with this idea, I expect larger changes in Internet usage and migration rates in communities closer to the cable network. Focussing first on the two plots on the left, Figure 3 depicts a striking relationship between the change in mean Internet usage and migration rates before and after the arrival of fast Internet in Nigeria and the distance to the cable network. Positive changes are more present in communities that are closer to the terrestrial cable network than in those further away. Interestingly, the relationship seems to be linear if we consider the distance in log instead of the absolute value in km, which motivates the use of log distance as a continuous measure of

[23]In fact, both approaches are equivalent if weights that account for the number of observations in a community and year are applied. The results of this paper do not change if I run the entire empirical analysis based on community times year level data. I refrain from this approach because using individual level data allows me to easily control for individual-specific covariates.

Figure 3:
Internet usage, migration, and distance to terrestrial cable network, 2010-12 change



Note: Binned scatter plot (10 equally sized bins) of difference in community mean Internet usage (top) and migration (bottom) between 2012 and 2010 and distance to terrestrial cable network in kilometres (left) and logartihmized distance to terrestrial cable network (right). 435 communities included. Figure A4 in the Appendix shows equivalent plots for Internet usage frequency.

the distance to the terrestrial cable network in the analysis below. Further, Figure A4 in the Appendix shows that the arrival of fast Internet did not only affect Internet take-up rates, but also the frequency with which individuals are using the Internet.

Table 2 reports the reduced form estimates of Equation (1). Each column of Table 2 reports the result of a separate regression of a binary variable indicating whether an individual migrated to another country and the log distance to the terrestrial cable network interacted with a year dummy for the post treatment period in the year 2012. In all regressions, I add community fixed effects as well as a dummy variable for the year 2012. In columns (2) and (4) I additionally control flexibly for a large set of control variables that is listed in Table 1. Columns (1) and (2) refer to the overall sample and columns (3) and (4) to a restricted sample including only individuals between the age of 20 and 35. For all specifications, Table 2 reports

Table 2:
Reduced form estimation: Migration on distance to terrestrial network

	(1)	(2)	(3)	(4)
Log(Distance to network) * Year 12	-0.0014***	-0.0014***	-0.0024***	-0.0024***
	(0.0005)	(0.0005)	(0.0009)	(0.0009)
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Huber-White (robust) standard errors clustered at the community level in parentheses, which allows for arbitrary correlation of the error term within a community to take into account serial correlation of the error term, which would otherwise severely downward-bias standard errors (Bertrand et al., 2004).

All four specifications of Table 2 show a positive effect of the arrival of fast Internet on migration. The interaction term reported in Table 2 indicates that communities close to the terrestrial cable network experienced a larger increase in migration rates that those farther away. The effect is virtually unaffected by the inclusion of control variables. For the entire sample, I estimate that a 100 % increase in the distance to the terrestrial cable network reduces the migration rate by around 0.14 percentage points or around 17.7 % relative to the sample mean in connected areas in the year 2012 (Distance < 5 km; mean: 0.79 %). The effect almost doubles in absolute terms if I restrict the sample to younger individuals, which supports the hypothesis that younger individuals are more responsive in Internet take-up to the availability of fast Internet, which then results in a higher change in the probability to migrate. A 100 % increase in the distance to the terrestrial cable network reduces the migration rate by around 0.24 percentage points. However, taking into consideration the higher likelihood of migrating for younger individuals irrespective of the exposure to faster Internet, the effect is surprisingly similar at around 16.5 % relative to the sample mean in connected areas in the year 2012 (mean: 1.45 %).

I provide a number of robustness and sensitivity tests in the Appendix. I use the absolute value in km instead of the logarithm of the distance and my results are qualitatively identical (Table A2). For this specification, my estimates suggests that an increase in the distance to the terrestrial cable network by 10 km reduces the change in the migration rate induced by the arrival of the submarine Internet cables by around 0.026 percentage points or around 10 % relative to the sample mean for the entire sample.

Table A3 reports results for a binary measure of distance. Communities are defined as being connected if they are within a 5 km radius around the terrestrial cable network. Panel A of Table A3 reports estimation results for the entire sample and Panel B for the restricted sample of younger individuals. In the first column of Table A3, I estimate that the availability of fast Internet lead to an increase in the migration rate by around 0.52 percentage points in communities close to the terrestrial cable network. In columns (2) to (4), I step-by-step exclude observations that are close to the connected area. If the point estimate changes by this procedure, this would indicate that the binary variable has been defined too narrowly. The downside of this procedure is that I possibly exclude individuals from locations that are more comparable to those in the treated areas. Nevertheless, it is comforting to see that the point estimate is hardly affected by the exclusion of these locations.

In Table A4, I exclude observations from remote communities. While there are arguments for including more remote locations in the sample as they are presumably less likely to be affected by the arrival of fast Internet than unconnected areas that are closer to the treated area, these locations might differ more from communities within the 5 km threshold around the terrestrial cable network. The estimates shown in Table A4 make clear that my findings are not driven by the inclusion of more remote areas. While some of the estimates become insignificant when excluding remote communities from my sample - likely due to the smaller sample size - the point estimates are remarkably unaffected.

A large part of the terrestrial cable network connects larger cities and runs through more urbanized areas. As a consequence, locations close to the terrestrial cable network and those further away differ in terms of a number of characteristics. For instance, as shown in Table A5, which reports mean values of selected community characteristics, locations close to the terrestrial cable network are more likely to be in the large cities Lagos or Abuja or, in general, in urban areas. Unsurprisingly, these locations differ also with respect to socio-economic characteristics of the individuals living in the community such as the share of individuals who are college educated or were Internet user already before the submarine cables arrived. While the fixed effect specification applied in the paper takes into account level differences between treated and untreated communities, I might wrongly attribute growth in migration rates to the arrival of fast Internet if locations in these more urbanized areas experienced faster growth in migration rates, irrespective of whether they are close to the terrestrial cable network or not. To see if this was the case, I construct placebo treatment variables by interacting the selected community characteristics listed in Table A5 with a dummy variable for the year 2012 which indicates the post-treatment period - similar to the treatment variable in Equation (1). Table A6 reports estimation results when these placebo treatment variables are added to the estimation equation. The estimated effect of the arrival of fast Internet is basically unchanged and the estimated coefficients on the placebo treatments are mostly not significant. An exemption is the placebo treatment for the large cities Lagos and Abuja in the restricted sample for younger individuals. However, since the estimated placebo effect is negative - which indicates that locations in these cities saw a smaller growth in migration rates between 2010 and 2012 - the baseline estimates presented above might even be underestimated.

A standard approach in the literature to assess the validity of difference-in-difference estimations - as applied in this paper - is to evaluate the trend in the outcome variable by treatment status for the pre-treatment period. If migration rates in connected and unconnected communities followed a similar trend before the arrival of fast Internet in Nigeria, I would expect migration rates to evolve similarly for both groups if the submarine Internet cables would not have been connected. Unfortunately, the Nigerian GHS data does not provide more than one pre-treatment period to apply such a check. Moreover, I am not aware of any other data set that provides information about geographical location and migration rates at the community level. However, to still be able to address concerns regarding the violation of the common trend assumption, I utilize information about remittances that remaining households receive. It is likely that migration rates in one period are correlated with the share of households that receive remittances in the following period. Hence, if migration rates in treated and untreated communities followed a similar trend before the arrival of fast Internet, I would expect that the difference in the share of households that receive remittances between households in connected and unconnected communities is similar in the years before the migration rates increased differently due to the connection of the submarine Internet cables. The time lag between migration and remittances allows me to evaluate the pre-treatment trend based on two periods as growth of migration rates in connected communities occurred only after the second wave in the year 2012. As a consequence, the difference in the share of households that receive remittances in connected and unconnected locations should not be different in the years 2010 and 2012. To assess if this is the case, Figure A5 shows an "event-study plot" for the differences in remittances between connected and unconnected locations by year. Indeed, as expected, I do not see a significant difference between the differences in remittances for locations close to the network and those further away in the year 2010 and 2012 but only in year 2015.

5.2 Instrumental variable estimates

Having established that the arrival of fast Internet in Nigeria lead to an significant increase in migration rates among locations close to the terrestrial cable network in comparison to more remote locations, I move on to investigate the impact on Internet usage on migration in the structural model explained in Section 4.2, exploiting the variation in Internet usage induced by the arrival of submarine Internet cables. Table 3 reports first stage and second stage estimates of the effect of Internet usage and Internet usage frequency on migration rates for the overall sample (columns (1) and (2)) and the restricted sample of younger individuals (columns (3) and (4)). Columns (2) and (4) refer to estimations where I additionally include the set of control variables listed in Table 1 of Section 3. The first two rows of Table 3 show the first stage es-

Table 3:
Instrumental variable estimation: Internet usage on migration

	(1)	(2)	(3)	(4)
First-stage estimates				
Internet usage				
Log(Distance to network) * Year 12	-0.014**	-0.014**	-0.023***	-0.021***
	(0.006)	(0.006)	(0.006)	(0.005)
F statistic	6.06	6.28	16.19	14.06
Internet usage frequency				
Log(Distance to network) * Year 12	-0.029***	-0.029***	-0.047***	-0.044***
	(0.009)	(0.009)	(0.010)	(0.010)
F statistic	9.72	10.26	21.58	18.89
Second-stage estimates				
Internet usage	0.096*	0.098*	0.105**	0.116**
	(0.056)	(0.056)	(0.048)	(0.054)
Internet usage frequency	0.047*	0.048**	0.050**	0.055**
	(0.024)	(0.024)	(0.022)	(0.024)
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

timates for the respective endogenous variable, i.e., the effect of the distance to the terrestrial cable network times the 2012 year dummy on Internet usage and the ordinal measure of Internet usage frequency, respectively. The third and forth row report the corresponding 2SLS estimate. In all specifications, I include a year fixed effect for the year 2012 as well as community fixed effects. Again, the reported robust standard errors in parentheses are clustered at the community level.

Focussing first on the first-stage estimates, Table 3 makes clear that the arrival of fast Internet lead to a significant change in the share of Internet users among communities that depends on the distance to the terrestrial cable network. As expected, locations closer to the terrestrial cable network saw a larger increase in the share of Internet users by the availability of fast Internet than more remote locations. This effect is significant for both, the entire sample as well as the restricted sample, and is not affected by the inclusion of control variables. My estimates indicate that an increase in the distance to the terrestrial cable network by 100 % reduces the share of Internet users by 1.4 percentage points for the entire sample and by 2.1 percentage points for younger individuals. The effect on the ordinal measure of Internet usage frequency is even larger. Here my estimates suggest that the arrival of fast Internet reduced the change in Internet usage frequency by around 2.9 or 4.4 percentage points of the ordinal measure of locations that are as twice as far away than other locations. Table 3 also reports the F-statistic on the reported instrument which is, in general, larger for the subsample of younger individuals and for the ordinal measure of Internet usage frequency, and in most instances above the rule-of-thumb-threshold of 10.

Turning next to the results of the second-stage estimation in the third and fourth row of Table 3, the reported results show a positive effect of Internet usage and Internet usage frequency on migration. Again, the estimates are hardly affected by the inclusion of socio-economic control variables. For the entire sample, I estimate that Internet usage increases the likelihood of migrating by around 9.8 percentage points. In terms of Internet usage frequency, my results suggest that an increase in Internet usage frequency by one increases the probability to migrate by 4.8 percentage points. Interestingly, the estimates do not differ for the unrestricted and restricted sample, which suggests that the effect is similar among older and younger individuals.

Again, I provide a number of sensitivity and robustness checks in the Appendix. Table A7 reports estimates when using a binary measure of distance to the terrestrial cable network instead of a continuous measure. As before, I define a location as being treated if it is within a 5 km radius around the terrestrial cable network. In the first column of Table A7, I use the entire sample for my estimations. Panel A refers to the full sample and Panel B shows the same estimates for the restricted sample of young individuals. Focussing first on the first-stage estimates in column (1), Table A7 indicates that the arrival of fast Internet lead to an increase in the share of Internet users by 5 percentage points among communities within the 5 km radius in comparison to those locations further away. To put this estimate into perspective, the share of individuals using the Internet in the year 2010 in the treated communities was at around 10 %, which suggests that the arrival of fast Internet increased Internet usage by 50 % among those individuals. Further, the corresponding second-stage estimates are remarkably similar to the baseline estimates in Table 3, which suggests that my baseline estimates are indeed driven by the boost in Internet usage among individuals located close to the terrestrial cable network. In columns (2) to (4), I step-by-step exclude observations that are close to the treated area as in the sensitivity analysis of the reduced form estimation. Again, I do not see large differences among the point estimates for the different samples, which confirms that the chosen 5 km corridor is not too narrow.

Table A8 reports estimates for the binary instrument when I exclude remote locations. Again, the idea is that locations far away from the treated communities are less comparable to the treated group. This might be of particular importance for Internet usage, where remote locations might lack any infrastructure that could help to access the Internet. It is remarkable that the point estimates of the first-stage effect do hardly differ when using the entire sample (column (1)) or the restricted sample that includes only observations within a 10 km radius around the terrestrial cable network. For this sample, the share of Internet users is very similar before the arrival of the submarine Internet cable (connected: 10 %, unconnected: 8 %; restricted: 13 % and 11 %). Again, this is very strong supporting evidence that the arrival of fast Internet was the driving force behind the increase in the share of Internet users.

In Table A9, I report results for the 2SLS when I additionally add the placebo treatment variables introduced above. The idea of this procedure is to allow for different growth paths among locations with particular characteristics more common in connected communities which results from the specific distribution of the terrestrial cable network in Nigeria. Focussing first on columns (1) to (5) where I step-by-step include separately the placebo-treatment variables, Table A9 makes clear that the largest effect of these placebo treatments can be found for binary variables indicating whether an individual resides in Lagos or Abuja or in urban areas. In both instances, I see a drop for the first-stage estimate, which increases the effect of Internet usage on migration, in particular for the Lagos or Abuja place treatment variable as the reduced form effect became even larger in this case. In all other cases, the second-stage effect remains robust to the inclusion of the placebo treatment. In sum, this suggests that my baseline estimates might even be on the lower bound of the effect of Internet usage on migration decisions.

Finally, I address concerns regarding the exclusion restriction of the excluded instrument in the instrumental variable estimation. As explained above, to interpret the second-stage estimate of Internet usage on migration as causal, the exclusion restriction for the excluded instrument needs to be fulfilled. This implies for my difference-in-difference set-up that there was no growth (or decline) in other factors between 2010 and 2012 that differed between connected and unconnected locations that might have impacted migration rates. Indeed, in a related study using a similar identification strategy, Hjort and Poulsen (2019) show that the arrival of the submarine Internet cables positively influenced employment rates in locations close to the terrestrial cable network in comparison to those further away. This might be problematic for my analysis as employment possibilities might influence migration decisions. For instance, better employment possibilities might reduce the incentives to migrate as individuals might be able to obtain a higher income level at their home location. On the contrary, finding employment might lower liquidity constraints that kept potential migrants from migrating. However, Hjort and Poulsen (2019) conduct their analysis based on different data and on a larger set of Sub-Saharan countries. Further, Hjort and Poulsen (2019) do not show how the results are driven by particular countries, in particular how important the arrival of the submarine Internet cable was in Nigeria. Hence, to alleviate concerns that I missattribute the change in migration rates to an increase in Internet usage instead of better employment perspectives, I investigate to what extend the arrival of fast Internet changed employment status among individuals in connected areas in Nigeria. Columns (1) and (2) in Table A10 in the

Appendix report reduced form estimates of Equation (1) when I use the employment status of an individual as dependent variable. The size and the signs of the obtained point estimates suggest that there was a positive effect of the arrival of fast Internet on employment status in Nigeria, too. However, the point estimates are estimated very imprecisely and are significantly different from zero only for the entire sample and the continuous measure of distance to terrestrial cable network. In columns (3) and (5) I add individuals' employment status to the baseline estimation equation and obtain estimates for the effect of the arrival of fast Internet that are essentially the same as in Table 3 and Table A7 in the Appendix. This is comforting as it shows that the interpretation of my estimates does not change if they are conditional on employment status. Next, I exclude all locations within a radius of 5 km of the terrestrial cable network that saw an increase in the mean employment rate between 2010 and 2012.[24] If my estimates are not affected by this restriction, this suggests that employment growth is not driving my results. Columns (4) and (6) of Table A10 show estimation results when excluding these observations. Remarkably, while the estimate loses precision for the binary measure of distance, the point estimates are hardly affected by the exclusion of these locations. In sum, I conclude that employment growth caused by the arrival of fast Internet is unlikely to be explaining my results.

Additionally, I check whether there was a change in other factors between 2010 and 2012 that differed between connected and unconnected locations that might potentially have impacted migration rates. In doing so, I run reduced from regressions for Equation (1) on the full set of control variables as listed in Table 1. The estimation results of this procedure are reported in Table A11. The first two columns of Table A11 refer to estimations when I use the continuous measure of distance and the last two columns to estimations when the binary measure of distance is included. Table A11 makes clear that there were hardly any systematic changes in the difference of these variables between connected and unconnected locations over time. Almost all estimates are insignificant or inconsistently estimated for the two measures of dis-

[24]To identify these locations, I first run a regression of employment status on a year dummy for the year 2012 and use the resulting residuals to calculate mean employment rates for each community. This allows me to abstract from any year specific effects on employment that affected all locations.

tance.[25] If anything, Table A11 suggests that there was a slight change from individuals with no schooling to some schooling and a reduction in mobile phone usage among individuals in connected areas in comparison to those in unconnected areas, with the latter suggesting that there might have been a substitution between between different types of ICTs. Overall, the results do not suggest that any of these factors might be explaining the change in migration rates.

6 Discussion

6.1 Migration out of Africa

As the previous section has shown, the arrival of fast Internet in Nigeria has caused an increase in migration rates, which was likely driven by increased Internet usage of individuals living close to the terrestrial cable network. In this subsection, I investigate whether increased migration due to Internet usage in Nigeria rather affected migration within the continent or migration out of Africa. An often overlooked fact is that a large share of international migration in Africa is within the continent. Recent estimates suggest that only roughly one-half of Africans who cross borders move out of the continent (International Organization for Migration, 2018). Nonetheless, the share of migrants living outside of Africa has experienced a sharp increase since 1990, when the same number was only at around 30 %. It seems plausible that Internet usage has stronger effects on migration out of Africa. First, if Internet usage affects migration behavior by providing potential migrants with information about destination countries which are not available for non-Internet users, I would expect the effect of Internet usage on migration to be larger if destination countries are (culturally) more remote. Similarly, if the exposure to foreign media transmitted via the Internet - which is dominated by images of Western lifestyle - changed individuals' preferences, I would also expect the effect of Internet usage to be larger for extra-continental migration.

[25] As for the baseline analysis, the estimated effect should be in different directions for the continuous measure and the binary measure.

Table 4:
Instrumental variable estimation: Migration out of Africa

	(1)	(2)	(3)	(4)	(5)	(6)
Internet usage	0.116**	0.070*	0.016			
	(0.054)	(0.039)	(0.013)			
Internet usage frequency				0.055**	0.033*	0.008
				(0.024)	(0.017)	(0.006)
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable:						
All migration	Yes	No	No	Yes	No	No
Migration out of Africa	No	Yes	No	No	Yes	No
Migration within in Africa	No	No	Yes	No	No	Yes
Observations	8,963	8,957	8,957	8,963	8,957	8,957
Cluster	435	435	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable is a binary variable indicating if an individual migrated to (1) another country, (2) out of Africa, and (3) within Africa (migration out of Africa is coded 0 in this case). Please note that the number of observations declines due to missing information about the destination country. Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4 shows Instrumental variable estimation results for different definitions of my dependent variable for the restricted sample of young individuals.[26] The first (second) row of Table 4 shows result for the effect of Internet usage (Internet usage frequency) on various measures of migration. Columns (1) and (4) show the baseline results from above for Internet usage and Internet usage frequency, respectively, where the dependent variable is a binary variable indicating whether an individual moved to another country. In columns (2) and (5), the dependent variable is equal to one if the individual moved to a country outside of Africa and zero otherwise, and column (3) and column (4) refer to estimation results when the dependent variable is equal to one if the individual moved to another country within Africa and zero otherwise. Note that I do not have information about the destination country for some migrants, so the number of observations differ between the first and second or third column

[26]Table A12 in the Appendix reports estimates for the entire sample. The results are qualitatively the same.

(forth and fifth or sixth column for Internet usage frequency).

Table 4 supports the idea that the effect of Internet usage on migration is larger for migration out of Africa. While I find positive effects of Internet usage on both outcome variables, the point estimate for migration within Africa is much smaller and not significant.

6.2 Internet usage, migration, and wealth

In this subsection, I provide estimation results to show whether the effect of Internet usage on migration depends on the wealth of potential migrants. Assuming that poverty is inversely related to skills (Angelucci, 2015), effect heterogeneity with respect to wealth implies that the spread of the Internet induces a change in the skill distribution of migrants. The self-selection and resulting skill distribution of immigrants is a major topic in the economic literature since Borjas (1987). While traditional approaches in the migration literature treated immigration as a change in the quantity of homogeneous labor supply, more recent work consider immigration within a framework of heterogeneous labor supply (e.g., Peri, 2016). A consequence of this approach is that the effect of immigration on the wages of natives depends on the skill distribution of immigrants and the elasticity of substitution across skill groups. Hence, it is important to know to what extent Internet usage facilitates migration of different skill groups. It is plausible to assume that the reduction in the costs of migrating due to Internet usage is larger for individuals from the lower part of the skill distribution. For instance, it is likely that the set of information that became available by using the Internet is larger for individuals from the lower part of the skill distribution, which suggest that responses due to Internet usage should also be larger for this group. Alternatively, studies have shown that migration from low-skilled individuals is often financially constrained (Chiquiar and Hanson, 2005; McKenzie and Rapoport, 2010; Bazzi, 2017). If access to the Internet reduces the cost of migrating, these previous constraints might be relaxed and low-skilled migration might become more common.

Table 5 reports estimates of the baseline 2SLS approach for subgroups defined by the relative household wealth for the sample of younger individuals.[27] I define individuals as being at

[27]Table A13 in the Appendix reports the same specifications for all individuals. Table A14 in the Appendix

the lower bound of the wealth distribution if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Column (1) reports results for the overall sample and column (2) and column (3) report estimation results for the two subsamples. As for the baseline estimates, Table 5 shows results for two measures of Internet usage - Internet up-take and Internet usage frequency.

Interestingly, the first-stage effects do hardly differ between the two subgroups. The reported F-statistic is slightly smaller for individuals at the lower bound of the wealth distribution, which might be the result of the smaller sample size, but the point estimates are very similar for low and high wealth individuals. However, the second-stage estimate differs considerably between both groups. While I do observe a positive effect of Internet usage for both subgroups, the effect is much smaller and not significantly different from zero for the sample of individuals with higher wealth. These results suggest that there is considerable effect heterogeneity of Internet usage on migration which might hint to a potential negative selection of migrants with respect to their skills.

7 Implications for remittances and economic development

The impact of international migration on the development in sending countries has received increasing attention in the literature. In particular, many scholars have argued that remittances sent by migrants are an important element for the well-being of the household members left behind. Rapoport and Docquier (2006) list a number or reasons why migrants send money back home, including altruism, repayment of loans to finance migration, or insurance and strategic motives. Such remittances contribute to the household income of the family members left behind and relax their liquidity constraints that might enable household members to purchase essential goods and escape poverty or to undertake investments in, e.g., businesses or children's education.[28]

reports results when the endogenous variable and the excluded instrument is interacted with a binary variable indicating the low and high household wealth. The results are qualitatively similar to the one presented here.

[28]The literature has examined the effect of remittances on a variety of outcomes. Common outcomes include income and measures of poverty, health, education, and asset ownership. For an overview, see, for instance, Gibson et al. (2011).

Table 5:
Instrumental variable estimation: Relative wealth

	(1)	(2)	(3)
First-stage estimates			
Internet usage			
Log(Distance to network) * Year 12	-0.021***	-0.018**	-0.021***
	(0.005)	(0.007)	(0.007)
F statistic	14.06	6.36	10.02
Internet usage frequency			
Log(Distance to network) * Year 12	-0.044***	-0.041***	-0.048***
	(0.010)	(0.016)	(0.015)
F statistic	18.89	6.77	10.06
Second-stage estimates			
Internet usage	0.116**	0.209**	0.038
	(0.054)	(0.105)	(0.033)
Internet usage frequency	0.055**	0.091**	0.017
	(0.024)	(0.046)	(0.015)
Year 12 FE	Yes	Yes	Yes
Community FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes
Restricted: Low wealth	No	Yes	No
Restricted: High wealth	No	No	Yes
Observations	8,963	3,925	5,038
Cluster	435	414	432

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

If the exposure to fast Internet and the resulting increase in Internet usage affected migration rates in Nigeria, I would also expect a different development of the share of households that receive remittance depending on the distance to the terrestrial cable network after the migration decision. Similarly, remaining household members might use remittances to invest in various outcomes which might, again, lead to different development paths between locations

Table 6:
Feedback effects: Migration and economic development

	(1)	(2)	(3)	(4)
$\mathbb{1}$(Distance to network $<$ 5km) * Year 15	0.029**	-0.077	0.022	0.093*
	(0.014)	(0.056)	(0.021)	(0.055)
$\mathbb{1}$(Distance to network $<$ 5km) * Year 10	0.003	0.027	0.009	-0.028
	(0.005)	(0.046)	(0.024)	(0.063)
Year FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Dependent variable:				
Remittances	Yes	No	No	No
Wealth items	No	Yes	No	No
Share HH member enrolled (age 10 to 18)	No	No	Yes	No
Share HH member enrolled (age 15 to 18)	No	No	No	Yes
Observations	10,414	10,414	8,031	3,196
Cluster	436	436	435	432

Note: Estimate of various outcome variables on interactions between a binary variable indicating whether a household is located within a 5 km radius around the terrestrial cable network and year dummies for the year 2010 and 2015. Number of observations is smaller in the third and fourth column as not all households have children in the depicted age bracket. Robust standard errors clustered at the community level in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

close to the terrestrial cable network and those more remote. I empirically analyze if such different growth paths are observable in my sample by estimating the following equation:

$$y_{h,c(h),t} = \mu_c + \sum_{j \in \{10,15\}} \{\beta_{0,j} \mathbb{1}\left[t = j\right] + \beta_{1,j} \mathbb{1}\left[t = j\right] * \mathbb{1}\left[Distance_{h,c(h)} < 5km\right]\} + u_{h,c(h),t}, \quad (4)$$

where $y_{h,c(h),t}$ is either a binary measure indicating whether household h in community $c(h)$ received remittances in year t or a measure of development. I estimate Equation (4) based on a sample at household level and include all three waves of the Nigerian GHS panel.

Table 6 reports estimation results of Equation (4) for various outcome variables. It is comforting to see that the interaction term is not significant before the arrival of fast Internet in Nigeria for all dependent variables. Further, the estimate shows that remittances increased sharply in connected areas relative to unconnected areas after the second migration period between the year 2012 and 2015. I do not find that these remittances are used to invest in wealth items. If anything, the effect of migration on wealth of families left behind might be even negative. One explanation could be that household members might have sold off assets to finance the migration costs of the migrant. Additionally, I do not find a significant effect on the share of household members that are enrolled for the age bracket 10 to 18. This might be due to compulsory schooling in Nigeria for children up to the age of 15. Focusing on children above the age of compulsory schooling, I do find a positive and significant effect.

Overall, the results suggest that the increase in migration due to the arrival of fast Internet had a positive effect on children's education in locations close to the terrestrial cable network. However, the causal channel might also be different. The availability of faster Internet might have a direct effect on remittances even without an increase in migration because it facilitates bank account transfers (Lee et al., 2020). Further, the positive effect on children's education might also be driven by changes in other factors due to arrival of fast Internet such as better educational infrastructure.

8 Conclusion

This paper provides evidence on the effect of the arrival of fast Internet on migration rates in Nigeria. Following Hjort and Poulsen (2019) I exploit the arrival of submarine Internet cables from Europe to Nigeria that increased Internet speed for individuals located close to the terrestrial cable network but not for others. Using this time and cross-sectional variation in a difference-in-difference approach, I show that locations close to the terrestrial cable network saw a larger increase in migration rates than remote locations after the arrival of fast Internet. Further, I provide evidence that this effect is likely driven by increased Internet usage among individuals located in communities close to the cable network. The effect of Internet usage on migration is more relevant for migration out of Africa. I further highlight interesting effect heterogeneity with respect to household wealth, suggesting that the spread of the Internet might lead to a negative selection of migrants. Finally, I show that the effect of the arrival of fast Internet on migration rates is followed by an increase in remittances among locations close to the terrestrial cable network. These remittances might be responsible for economic development in these locations, as they are also correlated with higher school enrolment.

References

Abadie, A. (2005): "Semiparametric difference-in-differences estimators," *Review of Economic Studies*, 72, 1–19.

Adetoro, N. (2010): "Internet utilisation and abuses in selected cybercafes in Ogun State, Nigeria," *African Journal of Library, Archives and Information Science*, 20, 19–28.

Adsera, A. and M. Pytlikova (2015): "The role of language in shaping international migration," *Economic Journal*, 125, 49–81.

Aker, J. C., M. A. Clemens, and C. Ksoll (2011): "Mobiles and mobility: The effect of mobile phones on migration in Niger," *Proceedings of the German Development Economics Conference, Berlin.*

Allen, T., C. d. C. Dobbin, and M. Morten (2018): "Border walls," *National Bureau of Economic Research*, No. 25267.

Angelucci, M. (2012): "US border enforcement and the net flow of Mexican illegal migration," *Economic Development and Cultural Change*, 60, 311–357.

——— (2015): "Migration and financial constraints: Evidence from Mexico," *Review of Economics and Statistics*, 97, 224–228.

BAMF (2019): "Aktuelle Zahlen, Oktober 2019," *Bundesamt für Migration und Flüchtlinge.*

Bazzi, S. (2017): "Wealth heterogeneity and the income elasticity of migration," *American Economic Journal: Applied Economics*, 9, 219–55.

Beine, M. and C. Parsons (2015): "Climatic factors as determinants of international migration," *Scandinavian Journal of Economics*, 117, 723–767.

Belot, M. and S. Ederveen (2012): "Cultural barriers in migration between OECD countries," *Journal of Population Economics*, 25, 1077–1105.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119, 249–275.

Black, R., S. Ammassari, S. Mouillesseaux, and R. Rajkotia (2004): "Migration and Pro-Poor Policy in West Africa, University of Sussex," *Development Research Centre on Migration, Globalisation and Poverty Sussex.*

Böhme, M. H., A. Gröger, and T. Stöhr (2020): "Searching for a better life: Predicting international migration with online search keywords," *Journal of Development Economics*, 142, 102347.

Borjas, G. J. (1987): "Self-Selection and the earnings of immigrants," *American Economic Review*, 531–553.

Braga, M. (2007): "Dreaming another life. The role of foreign media in migration decisions. Evidence from Albania," *Unpublished Working Paper.*

Brake, D. (2019): *Submarine Cables: Critical Infrastructure for Global Communications*, Information Technology & Innovation Foundation.

Cai, S. (2020): "Migration under liquidity constraints: Evidence from randomized credit access in China," *Journal of Development Economics*, 142.

Carrington, W. J., E. Detragiache, and T. Vishwanath (1996): "Migration with endogenous moving costs," *American Economic Review*, 909–930.

Cattaneo, C. and G. Peri (2016): "The migration response to increasing temperatures," *Journal of Development Economics*, 122, 127–146.

Chavula, J., N. Feamster, A. Bagula, and H. Suleman (2014): "Quantifying the effects of circuitous routes on the latency of intra-Africa internet traffic: a study of research and education networks," in *International Conference on e-Infrastructure and e-Services for Developing Countries*, Springer, 64–73.

Chiquiar, D. and G. H. Hanson (2005): "International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113, 239–281.

Clemens, M. A. (2014): "Does development reduce migration?" *International Handbook on Migration and Economic Development*, 152, 152–176.

Connor, P. (2018): *At Least a Million Sub-Saharan Africans Moved to Europe Since 2010: Sub-Saharan Migration to the United States also Growing*, Pew Research Center.

Czaika, M. and H. de Haas (2014): "The globalization of migration: Has the world become more migratory?" *International Migration Review*, 48, 283–323.

Dahl, G. and S. DellaVigna (2009): "Does movie violence increase violent crime?" *Quarterly Journal of Economics*, 124, 677–734.

De Chaisemartin, C. and X. d'Haultfoeuille (2018): "Fuzzy differences-in-differences," *Review of Economic Studies*, 85, 999–1028.

De Haas, H. (2006): *Trans-Saharan migration to North Africa and the EU: historical roots and current trends*, Migration Policy Institute, Migration Information Source.

——— (2007): "International migration, national development and the role of governments: The case of Nigeria," in *International Migration and National Development in Sub-Saharan Africa: viewpoints and policy initiatives in the country of origin*, Brill Publishers.

Dekker, R. and G. Engbersen (2014): "How social media transform migrant networks and facilitate migration," *Global Networks*, 14, 401–418.

DellaVigna, S. and E. La Ferrara (2015): "Economic and social impacts of the media," in *Handbook of Media Economics*, ed. by S. P. Anderson, J. Waldfogel, and D. Strömberg, Elsevier, vol. 1, 723–768.

Docquier, F. and H. Rapoport (2012): "Globalization, brain drain, and development," *Journal of Economic Literature*, 50, 681–730.

Dustmann, C. and A. Okatenko (2014): "Out-migration, wealth constraints, and the quality of local amenities," *Journal of Development Economics*, 110, 52–63.

Farré, L. and F. Fasani (2013): "Media exposure and internal migration: Evidence from Indonesia," *Journal of Development Economics*, 102, 48–61.

Feigenberg, B. (2020): "Fenced out: The impact of border construction on US-Mexico migration," *American Economic Journal: Applied Economics*, 12, 106–39.

Gentzkow, M. and J. M. Shapiro (2008): "Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study," *Quarterly Journal of Economics*, 123, 279–323.

Gibson, J., D. McKenzie, and S. Stillman (2011): "The impacts of international migration on remaining household members: omnibus results from a migration lottery program," *Review of Economics and Statistics*, 93, 1297–1318.

Greenstein, S. (2020): "The Basic Economics of Internet Infrastructure," *Journal of Economic Perspectives*, 34, 192–214.

GSMA (2015): "Digital Inclusion and the role of mobile in Nigeria," *www.gsma.com.*

Hanson, G. H. and C. McIntosh (2010): "The great Mexican emigration," *Review of Economics and Statistics*, 92, 798–810.

HATTON, T. J. (2005): "Explaining trends in UK immigration," *Journal of Population Economics*, 18, 719–740.

HERNÁNDEZ-COSS, R. AND C. EGWUAGU BUN (2007): *The UK-Nigeria remittance corridor: challenges of embracing formal transfer systems in a dual financial environment*, The World Bank.

HJORT, J. AND J. POULSEN (2019): "The arrival of fast internet and employment in Africa," *American Economic Review*, 109, 1032–79.

INTERNATIONAL ORGANIZATION FOR MIGRATION (2018): "World migration report 2018," Tech. rep.

KENDE, M. AND K. ROSE (2015): "Promoting local content hosting to develop the internet ecosystem," *www.internetsociety.org*.

LA FERRARA, E., A. CHONG, AND S. DURYEA (2012): "Soap operas and fertility: Evidence from Brazil," *American Economic Journal: Applied Economics*, 4, 1–31.

LEE, J. N., J. MORDUCH, S. RAVINDRAN, A. SHONCHOY, AND H. ZAMAN (2020): "Poverty and migration in the digital age: Experimental evidence on mobile banking in Bangladesh," *American Economic Journal: Applied Economics (forthcomming)*.

LU, Y., H. XIE, AND L. C. XU (2016): "Telecommunication externality on migration: Evidence from Chinese villages," *China Economic Review*, 39, 77–90.

MAYDA, A. M. (2010): "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23, 1249–1274.

MCKENZIE, D. AND H. RAPOPORT (2007): "Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico," *Journal of Development Economics*, 84, 1–24.

——— (2010): "Self-selection patterns in Mexico-US migration: the role of migration networks," *Review of Economics and Statistics*, 92, 811–821.

MISSIRIAN, A. AND W. SCHLENKER (2017): "Asylum applications respond to temperature fluctuations," *Science*, 358, 1610–1614.

MUNSHI, K. (2003): "Networks in the modern economy: Mexican migrants in the US labor market," *Quarterly Journal of Economics*, 118, 549–599.

MWESIGE, P. G. (2004): "Cyber elites: A survey of Internet café users in Uganda," *Telematics and Informatics*, 21, 83–101.

Ortega, F. and G. Peri (2013): "The effect of income and immigration policies on international migration," *Migration Studies*, 1, 47–74.

——— (2015): "Migration policies: recent advances on measurement, determinants and outcomes," *CESifo Economic Studies*, 61, 521–526.

Peri, G. (2016): "Immigrants, productivity, and labor markets," *Journal of Economic Perspectives*, 30, 3–30.

Rapoport, H. and F. Docquier (2006): "The economics of migrants' remittances," in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. by S.-C. Kolm and J. M. Ythier, Elsevier, vol. 2, 1135–1198.

Sjaastad, L. A. (1962): "The costs and returns of human migration," *Journal of Political Economy*, 70, 80–93.

Stanley, U., N. Nkordeh, B.-M. Ibiabo, and V. M. Olu (2018): "Broadband superhighway: A detailed analysis of Nigeria Trans-Atlantic fibre optic cables: Comparing MainOne cable, Glo-1 cable, WACS and SAT3," *Proceedings of the International Conference on Industrial Engineering and Operations Management*.

UN DESA, Population Division (2019): *World Population Prospects 2019, Volume II: Demographic Profiles (ST/ESA/SER.A/427)*, United Nations Department of Economic and Social Affairs.

UNDP (2019): *Human Development Report 2019. Beyond income, beyond averages, beyond today: Inequalitites in human development in the 21st century*, United Nations Development Programme.

WorldPovertyClock (2020): https://worldpoverty.io/, accessed: 2020-09-09.

Yang, D. (2011): "Migrant remittances," *Journal of Economic Perspectives*, 25, 129–52.

Young, A. (2013): "Inequality, the urban-rural gap, and migration," *Quarterly Journal of Economics*, 128, 1727–1785.

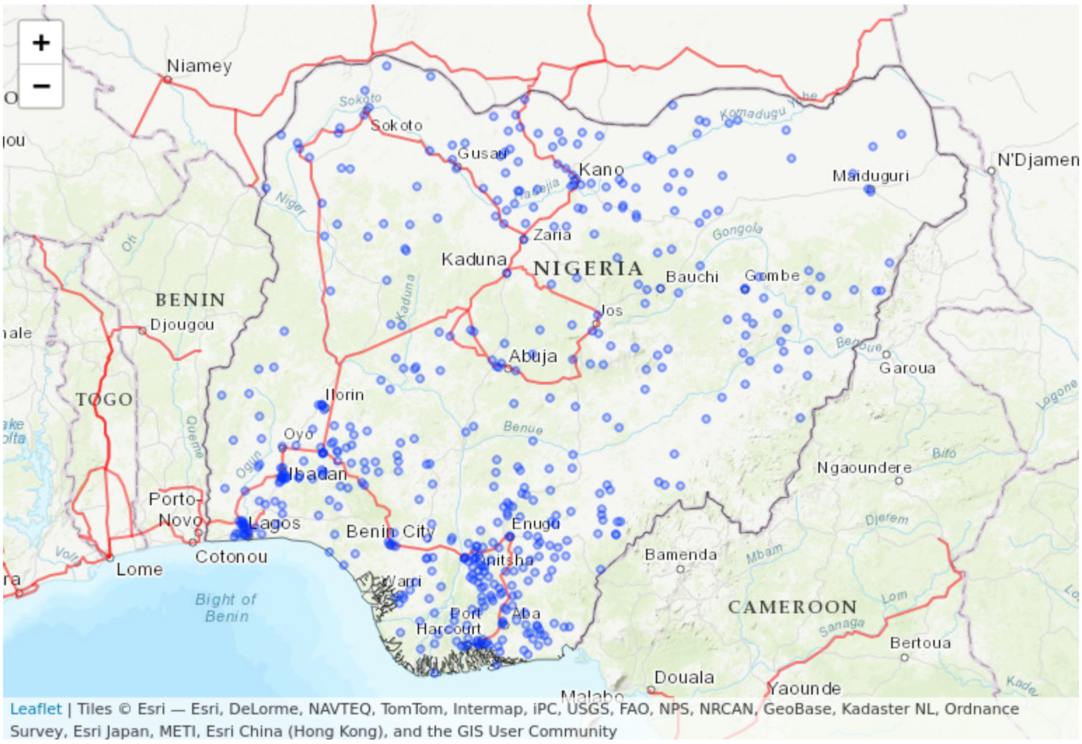
Zijlstra, J. and I. v. Liempt (2017): "Smart (phone) travelling: Understanding the use and impact of mobile technology on irregular migration journeys," *International Journal of Migration and Border Studies*, 3, 174–191.

Table A1:
OLS: International migration (binary) on Internet usage and Internet usage frequency

	(1)	(2)	(3)	(4)	(5)
Panel A					
Internet usage	0.0103***	0.0082**	0.0082**	0.0085**	0.0080**
	(0.0039)	(0.0039)	(0.0039)	(0.0038)	(0.0037)
Panel B					
Internet usage frequency	0.0051***	0.0041**	0.0042**	0.0044**	0.0043**
	(0.0019)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Year 12 FE	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes
State FE	No	No	Yes	No	No
County FE	No	No	No	Yes	No
Community FE	No	No	No	No	Yes
Observations	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435

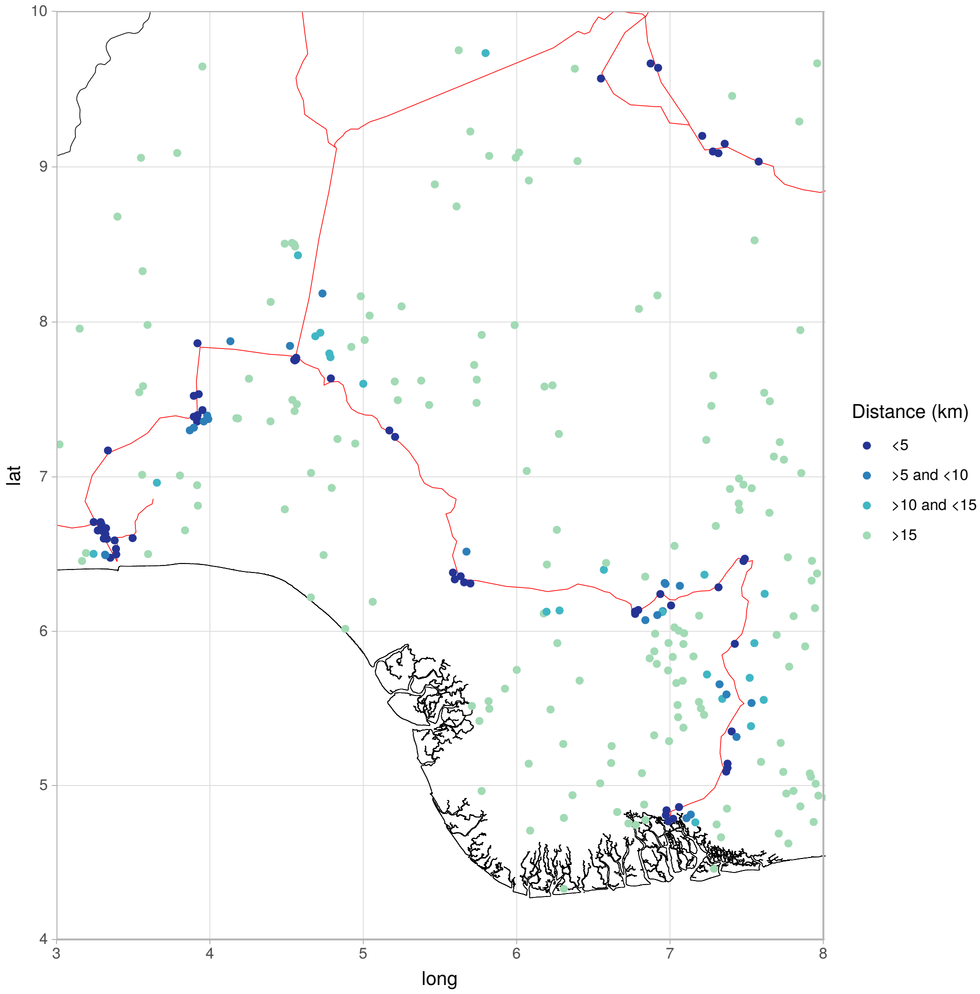
Note: Regression of a binary variable indicating whether an individual moved to another country on Internet usage and Internet usage frequency in the previous wave. Internet usage frequency is a ordinal measure (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1:
Nigeria, terrestrial cable network, and included communities



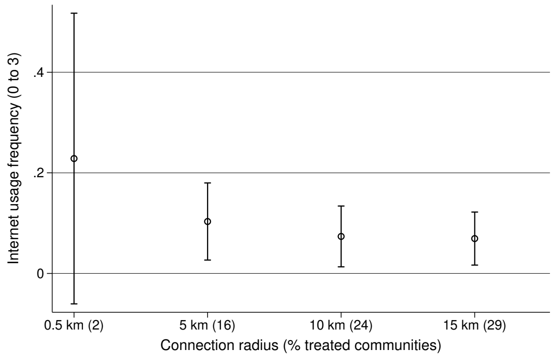
Notes: Red lines illustrate the diffusion of the terrestrial cable network in Nigeria and neighbouring countries. Blue dots indicate communities that are included in the final data set. Sources: Mapcruzin.com, Hjort and Poulsen (2019), Nigerian GHS panel.

Figure A2:
South Nigeria, terrestrial cable network, and included communities

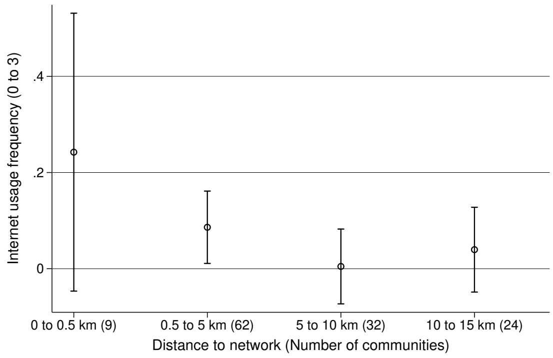


Notes: Red lines illustrate the diffusion of the terrestrial cable network in South Nigeria. Colored dots indicate communities included in the data set. Sources: Mapcruzin.com, Hjort and Poulsen (2019), Nigerian GHS panel.

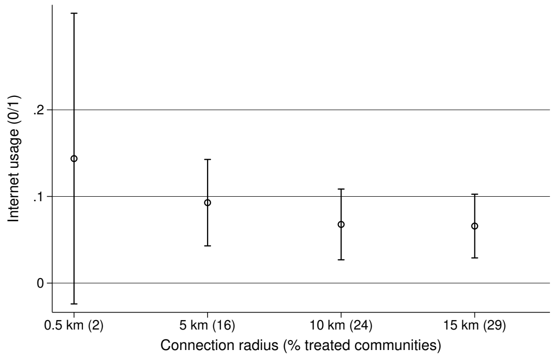
Figure A3:
Internet usage and distance to terrestrial cable network, 2010-12 change



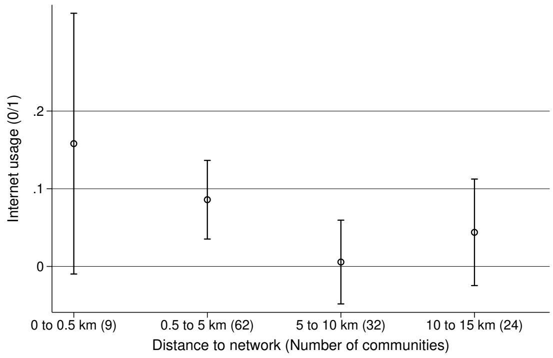
(a) Full sample, Internet usage frequency



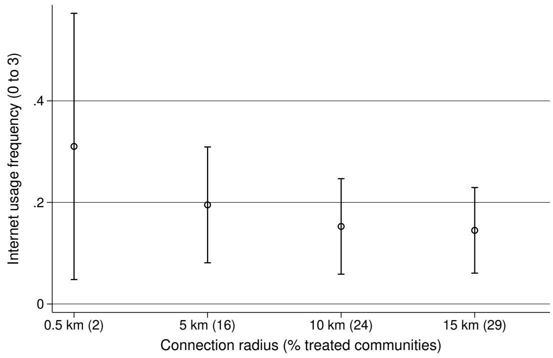
(b) Full sample, Internet usage frequency



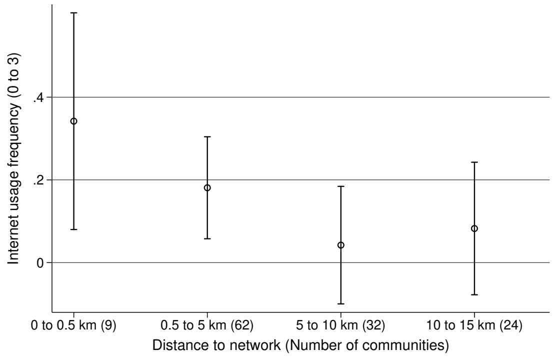
(c) Young individuals, Internet usage



(d) Young individuals, Internet usage



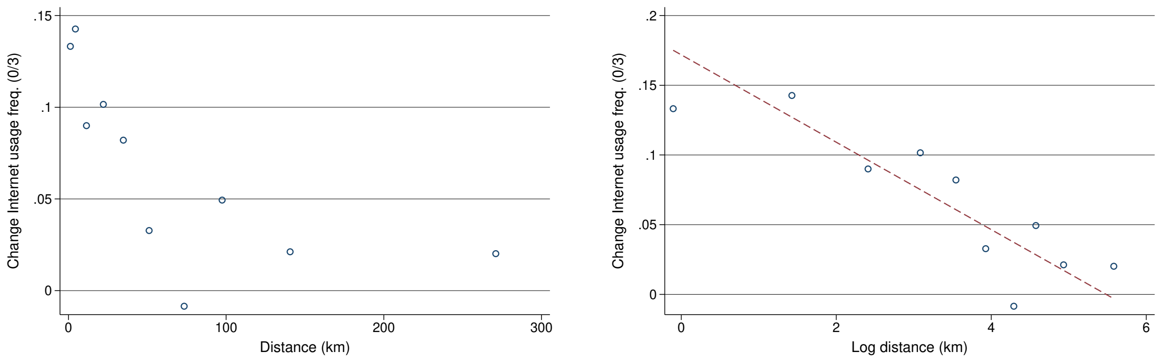
(e) Young individuals, Internet usage frequency



(f) Young individuals, Internet usage frequency

Note: Plots on the left show coefficient estimates for four separate regressions of Internet usage or Internet usage frequency on an interaction term of a binary variable indicating if individual i is located in a community within the connection radius shown on the x-axis to the terrestrial cable network times an indicator variable for the year 2012. Plots on the right show coefficient estimates for a regression of Internet usage or Internet usage frequency on a set of binary variables indicating if individual i is located within a bin shown on the x-axis (baseline: Distance to terrestrial cable network larger than 15 km). All estimates include a year dummy for the year 2012 as well as community fixed effects. Young individuals are between 20 and 35 at interview date. Number of observations: 21,626 (full sample), 8,963 (young individuals). 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster).

Figure A4:
Internet usage frequency and distance to terrestrial cable network, 2010-12 change



Note: Binned scatter plot (10 equally sized bins) of difference in community mean Internet usage frequency between 2012 and 2010 and distance to terrestrial cable network in kilometres (left) and logartihmized distance to terrestrial cable network (right). 435 communities included.

Table A2:
Reduced form estimation: Robustness, network distance not logarithmized

	(1)	(2)	(3)	(4)
Distance to network * Year 12	-0.0026***	-0.0027***	-0.0037***	-0.0037***
	(0.0008)	(0.0008)	(0.0012)	(0.0012)
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Restricted: Age 20 to 35	No	No	Yes	Yes
Observations	21,626	21,626	8,963	8,963
Cluster	435	435	435	435

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3:
Reduced form estimation: Robustness, binary measure

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0052*	0.0053*	0.0053*	0.0062**
	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Observations	21,626	20,259	18,406	16,721
Cluster	435	403	361	327
Panel B: Age 20 to 35				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0104*	0.0106*	0.0108*	0.0117**
	(0.0055)	(0.0055)	(0.0055)	(0.0055)
Observations	8,963	8,408	7,682	6,986
Cluster	435	403	361	327
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance between 5 and 10 km	No	Yes	Yes	Yes
Distance between 10 and 20 km	No	No	Yes	Yes
Distance between 20 to 30 km	No	No	No	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4:
Reduced form estimation: Robustness, binary measure, excluding remote locations

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0052*	0.0041	0.0046	0.0049
	(0.0029)	(0.0030)	(0.0034)	(0.0048)
Observations	21,626	15,944	6,492	4,639
Cluster	435	333	145	103
Panel B: Age 20 to 35				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0104*	0.0093*	0.0085	0.0085
	(0.0055)	(0.0056)	(0.0063)	(0.0080)
Observations	8,963	6,582	2,696	1,970
Cluster	435	333	145	103
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance $>$ 100 km	No	Yes	Yes	Yes
Distance $>$ 20 km	No	No	Yes	Yes
Distance $>$ 10 km	No	No	No	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5:
Community mean values by distance to the terrestrial cable network

	Distance $> 5\,\text{km}$	Distance $< 5\,\text{km}$	Mean diff.
Community			
located in states: Lagos, Abuja	0.02	0.17	-0.15***
			(0.03)
located in urban area	0.22	0.73	-0.52***
			(0.05)
with high share of educated individuals in 2010	0.23	0.38	-0.15**
			(0.06)
with high share of Internet user in 2010	0.21	0.49	-0.28***
			(0.06)

Note: Mean values and mean difference tests of selected community characteristics by distance to the terrestrial cable network. A community has a high share of Internet users (educated individuals) if the share of Internet users (college educated individuals) is in the highest quartile in the sample. Sample size: 435 (distance < 5 km: 71). Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6:
Reduced form estimation: Robustness, additional controls

	(1)	(2)	(3)	(4)	(5)
Panel A: Entire sample					
Log(Distance to network) * Year 12	-0.0015***	-0.0011**	-0.0011***	-0.0013***	-0.0011***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0004)
Lagos or Abuja * Year 12	-0.0034				-0.0044
	(0.0036)				(0.0038)
Urban * Year 12		0.0026			0.0016
		(0.0020)			(0.0021)
Internet usage year 10 * Year 12			0.0037*		0.0033
			(0.0021)		(0.0026)
Education year 10 * Year 12				0.0014	-0.0003
				(0.0023)	(0.0029)
Observations	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435
Panel B: Age 20 to 35					
Log(Distance to network) * Year 12	-0.0027***	-0.0021**	-0.0021***	-0.0021**	-0.0022***
	(0.0010)	(0.0008)	(0.0007)	(0.0008)	(0.0008)
Lagos or Abuja * Year 12	-0.0103***				-0.0124***
	(0.0039)				(0.0048)
Urban * Year 12		0.0030			0.0024
		(0.0038)			(0.0039)
Internet usage year 10 * Year 12			0.0040		-0.0001
			(0.0038)		(0.0045)
Education year 10 * Year 12				0.0064	0.0064
				(0.0041)	(0.0047)
Observations	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435
Included covariates (Panel A and B):					
Year 12 FE	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). *Lagos or Abuja* is a binary variable indicating whether an individual is located either in Logos or Abuja, *Urban* is a binary variable indicating whether an individual resides in an urban area. *Internet usage year 10* is a binary variable indicating whether an individual lives in a community where the share of Internet users in 2010 was in the highest quartile in the sample. *Education year 10* is a binary variable indicating whether an individual lives in a community where the share of college educated individuals in 2010 was in the highest quartile in the sample. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A5:
Robustness: Parallel pre-trends, remittances

Note: Plot shows estimated coefficients $\beta_{1,10}$ and $\beta_{1,15}$ of the equation:

$$Remittances_{h,c(h),t} = \mu_c + \sum_{j\in\{10,15\}} \{\beta_{0,j}\mathbb{1}\,[t=j] + \beta_{1,j}\mathbb{1}\,[t=j] * \mathbb{1}\left[Distance_{h,c(h)} < 5km\right]\} + \epsilon_{h,c(h),t},$$

where $Remittances_{h,c(h),t}$ is a binary variable indicating whether household h located in community $c(h)$ has received remittances within 12 month before the interview year t, and μ_c represents a set of community fixed effects. 2010 (2012, 2015) refers to the first (second, third) wave of the Nigerian GHS panel. Estimates are based on a sample at household level. Number of observations: 10,414. Plotted 95 % confidence intervals are based on cluster-robust standard errors at the community level (435 cluster). Point estimates and standard errors can be found in column (1) of Table 6 in Section 7.

Table A7:
Instrumental variable estimation: Robustness, binary instrument

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
First-stage estimates				
Internet usage				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.050**	0.050**	0.051**	0.053***
	(0.020)	(0.020)	(0.020)	(0.020)
F statistic	6.24	6.22	6.52	7.01
Internet usage frequency				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.105***	0.106***	0.108***	0.113***
	(0.039)	(0.039)	(0.039)	(0.039)
F statistic	7.23	7.29	7.63	8.37
Second-stage estimates				
Internet usage	0.104	0.106	0.101	0.101
	(0.072)	(0.073)	(0.070)	(0.068)
Internet usage frequency	0.049	0.050	0.047	0.047
	(0.033)	(0.033)	(0.032)	(0.031)
Observations	21,626	20,259	19,202	18,406
Cluster	435	403	379	361
Panel B: Restricted: Age 20 to 35				
First-stage estimates				
Internet usage				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.092***	0.092***	0.094***	0.096***
	(0.026)	(0.026)	(0.026)	(0.026)
F statistic	12.40	12.27	12.99	13.38
Internet usage frequency				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.194***	0.196***	0.200***	0.206***
	(0.058)	(0.058)	(0.058)	(0.058)
F statistic	11.15	11.36	11.94	12.72
Second-stage estimates				
Internet usage	0.112*	0.115*	0.111*	0.112*
	(0.062)	(0.063)	(0.061)	(0.060)
Internet usage frequency	0.053*	0.054*	0.052*	0.052*
	(0.031)	(0.031)	(0.030)	(0.029)
Observations	8,963	8,408	8,014	7,682
Cluster	435	403	379	361
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance between 5 and 10 km	No	Yes	Yes	Yes
Distance between 10 and 15 km	No	No	Yes	Yes
Distance between 15 and 20 km	No	No	No	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Binary variable indicating if distance to the terrestrial cable network is below 5 km times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are the same as for the baseline estimates (Table 3). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8:
Instrumental variable estimation: Robustness, binary instrument, excluding remote

	(1)	(2)	(3)	(4)
Panel A: Entire sample				
First-stage estimates				
Internet usage				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.050**	0.042**	0.035	0.050*
	(0.020)	(0.020)	(0.023)	(0.027)
F statistic	6.24	4.46	2.43	3.63
Internet usage frequency				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.105***	0.091**	0.068	0.098*
	(0.039)	(0.039)	(0.046)	(0.055)
F statistic	7.23	5.25	2.16	3.23
Second-stage estimates				
Internet usage	0.104	0.097	0.132	0.097
	(0.072)	(0.084)	(0.127)	(0.105)
Internet usage frequency	0.049	0.045	0.068	0.050
	(0.033)	(0.039)	(0.068)	(0.055)
Observations	21,626	15,944	6,492	4,639
Cluster	435	333	145	103
Panel B: Restricted: Age 20 to 35				
First-stage estimates				
Internet usage				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.092***	0.082***	0.075**	0.100***
	(0.026)	(0.027)	(0.031)	(0.036)
F statistic	12.40	9.47	5.77	7.48
Internet usage frequency				
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.194***	0.175***	0.136*	0.179**
	(0.058)	(0.059)	(0.072)	(0.088)
F statistic	11.15	8.69	3.55	4.13
Second-stage estimates				
Internet usage	0.112*	0.113	0.112	0.085
	(0.062)	(0.071)	(0.088)	(0.084)
Internet usage frequency	0.053*	0.053	0.062	0.047
	(0.031)	(0.035)	(0.054)	(0.050)
Observations	8,963	6,582	2,696	1,970
Cluster	435	333	145	103
Included covariates (Panel A and B):				
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Excluded observations (Panel A and B):				
Distance $>$ 100 km	No	Yes	Yes	Yes
Distance $>$ 15km	No	No	Yes	Yes
Distance $>$ 10km	No	No	No	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Binary variable indicating if distance to the terrestrial cable network is below 5 km times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Control variables included are the same as for the baseline estimates (Table 3). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A9:
Instrumental variable estimation: Robustness, additional controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Entire sample						
First-stage estimates						
Internet usage						
Log(Distance to network) * Year 12	-0.014**	-0.010**	-0.012**	-0.015**	-0.014**	-0.009**
	(0.006)	(0.004)	(0.006)	(0.006)	(0.005)	(0.004)
F statistic	6.28	5.74	4.08	6.25	6.11	4.22
Internet usage frequency						
Log(Distance to network) * Year 12	-0.029***	-0.019***	-0.022**	-0.028***	-0.027***	-0.014**
	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)	(0.007)
F statistic	10.26	7.73	5.39	8.59	9.47	3.94
Second-stage estimates						
Internet usage	0.098*	0.150*	0.094	0.074*	0.096*	0.124
	(0.056)	(0.084)	(0.065)	(0.043)	(0.055)	(0.079)
Internet usage frequency	0.048**	0.078*	0.051	0.040*	0.049**	0.079
	(0.024)	(0.040)	(0.033)	(0.021)	(0.025)	(0.051)
Observations	21,626	21,626	21,626	21,626	21,626	21,626
Cluster	435	435	435	435	435	435
Panel B: Restricted: Age 20 to 35						
First-stage estimates						
Internet usage						
Log(Distance to network) * Year 12	-0.021***	-0.016***	-0.016***	-0.021***	-0.019***	-0.013***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
F statistic	14.06	10.40	8.47	13.32	13.17	7.23
Internet usage frequency						
Log(Distance to network) * Year 12	-0.044***	-0.031***	-0.031***	-0.043***	-0.040***	-0.022**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
F statistic	18.89	10.74	10.07	17.76	17.37	5.92
Second-stage estimates						
Internet usage	0.116**	0.174**	0.127*	0.099**	0.106**	0.171**
	(0.054)	(0.079)	(0.067)	(0.044)	(0.051)	(0.084)
Internet usage frequency	0.055**	0.087**	0.067**	0.050**	0.052**	0.102*
	(0.024)	(0.040)	(0.034)	(0.021)	(0.024)	(0.052)
Observations	8,963	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435	435
Included covariates (Panel A and B):						
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	No
Lagos or Abuja * Year 12	No	Yes	No	No	No	Yes
Urban * Year 12	No	No	Yes	No	No	Yes
Internet usage year 10 * Year 12	No	No	No	Yes	No	Yes
Education year 10 * Year 12	No	No	No	No	Yes	Yes

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Control variables included are the same as for the baseline estimates (Table 3). *Lagos or Abuja* is a binary variable indicating whether an individual is located either in Logos or Abuja. *Urban* is a binary variable indicating whether an individual resides in an urban area. *Internet usage year 10* is a binary variable indicating whether an individual lives in a community where the share of Internet users in 2010 was in the highest quartile in the sample. *Education year 10* is a binary variable indicating whether an individual lives in a community where the share of college educated individuals in 2010 was in the highest quartile in the sample. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A10:
Reduced form estimation: Robustness, exclusion restriction

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Entire sample						
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0079		0.0052*	0.0060		
	(0.0182)		(0.0029)	(0.0042)		
Log(Distance to network) * Year 12		-0.0083**			-0.0014***	-0.0017***
		(0.0037)			(0.0005)	(0.0006)
Observations	21,626	21,626	21,626	20,391	21,626	20,391
Cluster	435	435	435	406	435	406
Panel B: Age 20 to 35						
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	0.0268		0.0105*	0.0117		
	(0.0279)		(0.0055)	(0.0078)		
Log(Distance to network) * Year 12		-0.0100			-0.0024***	-0.0027**
		(0.0061)			(0.0009)	(0.0012)
Observations	8,963	8,963	8,963	8,442	8,963	8,442
Cluster	435	435	435	406	435	406
Included covariates (Panel A and B):						
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional control: Employment status	No	No	Yes	No	Yes	No
Excluded communities (Panel A and B):						
Distance $<$ 5 km & Empl. growth $>$ 0	No	No	No	Yes	No	Yes
Dependent variable (Panel A and B):						
Employment status	Yes	Yes	No	No	No	No
Migration	No	No	Yes	Yes	Yes	Yes

Note: Dependent variable is either a binary variable indicating whether an individual was employed within the last 7 days or a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A11:
Reduced form estimation: Robustness, exclusion restriction other

	(1)	(2)	(3)	(4)
Dependent variable:				
Age	0.0452	-0.0100	-0.0127	0.2536
	(0.0527)	(0.0455)	(0.2324)	(0.1728)
Female	-0.0006	-0.0015	0.0019	0.0062
	(0.0018)	(0.0029)	(0.0089)	(0.0152)
Household head	-0.0005	-0.0040	-0.0032	0.0039
	(0.0013)	(0.0025)	(0.0064)	(0.0117)
Spouse	-0.0001	-0.0017	0.0009	0.0037
	(0.0016)	(0.0027)	(0.0081)	(0.0129)
Son/Daughter	0.0016	0.0027	0.0034	0.0144
	(0.0021)	(0.0038)	(0.0105)	(0.0180)
Other household member	-0.0010	0.0030	-0.0011	-0.0220**
	(0.0012)	(0.0021)	(0.0053)	(0.0087)
Relation to HH head (ordinal)	0.0001	0.0126**	0.0045	-0.0335
	(0.0035)	(0.0063)	(0.0163)	(0.0279)
No schooling	0.0052	0.0061	-0.0365	-0.0339
	(0.0043)	(0.0044)	(0.0246)	(0.0217)
Some schooling	-0.0105**	-0.0118**	0.0391	0.0435
	(0.0048)	(0.0054)	(0.0264)	(0.0282)
Secondary education	0.0057	0.0093	-0.0088	-0.0182
	(0.0035)	(0.0059)	(0.0164)	(0.0256)
University degree	-0.0003	-0.0037	0.0062	0.0086
	(0.0017)	(0.0033)	(0.0075)	(0.0147)
Education (ordinal)	-0.0002	-0.0042	0.0401	0.0329
	(0.0069)	(0.0087)	(0.0326)	(0.0365)
Wealth items = 0	-0.0029	-0.0029	0.0290**	0.0162
	(0.0044)	(0.0050)	(0.0143)	(0.0162)
Wealth items = 1	0.0070	0.0101	-0.0389	-0.0352
	(0.0071)	(0.0078)	(0.0291)	(0.0322)
Wealth items = 2	-0.0080	-0.0074	0.0354	0.0240
	(0.0066)	(0.0076)	(0.0313)	(0.0357)
Wealth items = 3	0.0039	-0.0006	-0.0346	-0.0052
	(0.0058)	(0.0072)	(0.0291)	(0.0367)
Wealth items = 4	0.0000	0.0008	0.0090	0.0003
	(0.0030)	(0.0034)	(0.0182)	(0.0220)
Wealth items (ordinal)	0.0027	-0.0033	-0.0356	-0.0019
	(0.0109)	(0.0130)	(0.0509)	(0.0584)
Enrolled	-0.0019	-0.0068	-0.0017	0.0025
	(0.0034)	(0.0051)	(0.0162)	(0.0239)
TV usage	0.0013	0.0081	0.0053	-0.0057
	(0.0051)	(0.0059)	(0.0204)	(0.0222)
Mobile phone usage	0.0119**	0.0103*	-0.0081	0.0002
	(0.0060)	(0.0061)	(0.0268)	(0.0271)
Reported coefficient:				
Log(Distance to network) * Year 12	Yes	Yes	No	No
$\mathbb{1}$(Distance to network $<$ 5km) * Year 12	No	No	Yes	Yes
Year 12 FE	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	No	Yes	No	Yes
Observations	21,626	8,963	21,626	8,963
Cluster	435	435	435	435

Note: First column specifies the dependent variable of a regression on measures of distance to the terrestrial cable network times an indicator variable for the year 2012. Control variables included are listed in Table 1 (dependent variable excluded). Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A12:
Instrumental variable estimation: Migration out of Africa, full sample

	(1)	(2)	(3)	(4)	(5)	(6)
Internet usage	0.098*	0.049	0.016			
	(0.056)	(0.032)	(0.013)			
Internet usage frequency				0.048**	0.024*	0.008
				(0.024)	(0.014)	(0.006)
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable:						
All migration	Yes	No	No	Yes	No	No
Migration out of Africa	No	Yes	No	No	Yes	No
Migration within in Africa	No	No	Yes	No	No	Yes
Observations	21,626	21,612	21,612	21,626	21,612	21,612
Cluster	435	435	435	435	435	435

Note: Dependent variable is a binary variable indicating if an individual migrated to another country. Control variables included are: Age, sex (binary), household member (binary: head, spouse, son/daughter, other), enrolled in school (binary), highest education (binary: no schooling, some schooling, secondary education, university degree), number of wealth items (binary: 0 to 4), other ICT usage (binary: mobile phone, TV). *Lagos or Abuja* is a binary variable indicating whether an individual is located either in Logos or Abuja, *Urban* is a binary variable indicating whether an individual resides in an urban area. *Internet usage year 10* is a binary variable indicating whether an individual lives in a community where the share of Internet users in 2010 was in the highest quartile in the sample. *Education year 10* is a binary variable indicating whether an individual lives in a community where the share of college educated individuals in 2010 was in the highest quartile in the sample. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A13:
Instrumental variable estimation: Relative wealth, full sample

	(1)	(2)	(3)
First-stage estimates			
Internet usage			
Log(Distance to network) * Year 12	-0.014**	-0.008**	-0.012**
	(0.006)	(0.004)	(0.005)
F statistic	6.28	4.28	5.63
Internet usage frequency			
Log(Distance to network) * Year 12	-0.029***	-0.018**	-0.029***
	(0.009)	(0.008)	(0.011)
F statistic	10.26	4.82	7.28
Second-stage estimates			
Internet usage	0.098*	0.273	0.032
	(0.056)	(0.175)	(0.032)
Internet usage frequency	0.048**	0.120	0.014
	(0.024)	(0.073)	(0.013)
Year 12 FE	Yes	Yes	Yes
Community FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Restricted: Low wealth	No	Yes	No
Restricted: High wealth	No	No	Yes
Observations	21,626	9,733	11,893
Cluster	435	421	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage frequency on migration decisions. Excluded instrument: Log distance to the terrestrial cable network times an indicator variable for the year 2012. Dependent variable of the first-stage estimates in the first (second) row is Internet usage (Internet usage frequency). Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A14:
Instrumental variable estimation: Relative wealth, 2 endogenous variables

	(1)	(2)	(3)	(4)	(5)	(6)
First-stage estimate						
Log(Dist. to network) * Year 12	-0.021***	0.001	-0.045***	0.003		
	(0.007)	(0.002)	(0.014)	(0.004)		
Log(Dist. to network) * Year 12 * Low wealth	0.003	-0.019***	0.005	-0.044***		
	(0.010)	(0.006)	(0.021)	(0.014)		
Second-stage estimates						
Internet usage * Low wealth					0.149	
					(0.104)	
Internet usage					0.068	
					(0.050)	
Internet usage frequency * Low wealth						0.064
						(0.045)
Internet usage frequency						0.032
						(0.022)
F-statistic (First stage)					10.80	11.04
Endogenous variable:						
Internet usage	Yes	Yes	No	No	Yes	No
Internet usage frequency	No	No	Yes	Yes	No	Yes
Year 12 FE	Yes	Yes	Yes	Yes	Yes	Yes
Community FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Restricted: Age 20 to 35	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,963	8,963	8,963	8,963	8,963	8,963
Cluster	435	435	435	435	435	435

Note: Instrumental variable estimates of the effect of Internet usage and Internet usage times a binary variable indicating low wealth and Internet usage frequency and Internet usage frequency times an binary variable indicating low wealth on migration decisions. Excluded instruments: Log distance to the terrestrial cable network times an indicator variable for the year 2012 and an interaction with a binary variable indicating low wealth. Internet usage is a binary variable indicating whether an individual reported in the survey interview that he or she has access to the Internet. Internet usage frequency is an ordinal measure of frequency (0 = less than a month / no access, 1 = at least once a month, 2 = at least once a week, 3 = daily). Dependent variable of the second-stage estimates in the third and fourth rows is a binary variable indicating if an individual migrated to another country. Individuals are defined as having *low wealth* if the number of wealth items of their household is below the mean of the number of wealth items in the respective community in which they are living. Individuals with *high wealth* are all other individuals. Control variables included are listed in Table 1. Robust standard errors clustered at the community level in parentheses.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.