Digital Infrastructure and Local Economic Growth: Early Internet in Sub-Saharan Africa*

Moritz Goldbeck† Valentin Lindlacher‡

April 12, 2021

PRELIMINARY AND INCOMPLETE
Please do not cite nor distribute.

Abstract

Digital infrastructure enables widespread access to modern information and communication technologies—most prominently the Internet—promising to stimulate economic growth. We analyze Sub-Saharan African (SSA) countries in the early 2000s and ask if Internet availability even at basic speeds contributes to regional economic development. Exploiting quasi-random variation in Internet availability induced by sub-marine cable arrivals in a difference-in-differences setting, we measure the growth of SSA cities using nighttime light data. Our findings suggest that Internet availability at basic speeds leads to three percentage points higher economic growth of SSA cities in the years after connection compared to a control group of similar cities not (yet) connected.

Keywords: Internet, regional/urban development, cities, nighttime light

JEL-Codes: O33, O18, R11

*We thank Lena Abou El-Komboz, Moritz Dillmann, Nicolas Göller, and Leon Wolf for excellent research assistance. Further, we thank Vojtech Bartos, Mathias Bühler, Oliver Falck, Lukasz Grzybowski, Anna Kerkhof, Markus Ludwig, Onkokame Mothobi, Niklas Potrafke, Helmut Rainer and Maria Waldinger for valuable comments and suggestions. Finally, we thank the participants of the (internal) seminars at the ifo Institute and at Economic Research South Africa. All errors are our own.

†University of Munich & ifo Institute, goldbeck@ifo.de.

‡University of Munich & ifo Institute, lindlacher@ifo.de.

1 Introduction

In the last decades, the provision of digital infrastructure in many countries enabled widespread access and adoption of modern information and communication technologies (ICT), most prominently the Internet. Evidence shows positive effects of (broadband) Internet availability on individual-level economic performance (Akerman et al., 2015) and country-level economic growth (Czernich et al., 2011) for developed countries. Hopes are high that Internet access can foster regional economic growth in the developing world as well (World Bank, 2016). In Sub-Saharan Africa (SSA) for example—where impulses for economic growth are required to fight poverty and deprivation—Governments, public-private partnerships, and companies alike invest large amounts of money to bring the Internet to everyone.[1]

Despite these enormous investments, a growth effect of Internet in SSA is less obvious than it seems. On the one hand, lacking legacy infrastructure (i.e., fixed-line telephony networks) to build on makes the provision more complex and costly. At the same time, low population density apart from a few mega-cities, missing hardware, financial constraints and a lower willingness to pay lead to lower adoption rates (World Bank, 2016). On the other hand, the potential of Internet seems particularly high in SSA since alternative ICT like fixed-line telephony is largely absent. It is thus crucial to investigate the effect of Internet availability on regional economic development in a developing-country context. With the notable exception of Hjort and Poulsen (2019), who find sizable positive individual-level effects of broadband Internet on employment in SSA around 2010, evidence on the economic impact of Internet availability in developing countries is surprisingly rare.

In this paper, we ask if there is a causal effect of availability of Internet even at basic speeds on local economic growth in SSA. We focus on the very first wave of Internet available in SSA since the early 2000s, where only basic Internet speed was available in some countries. To investigate if potential individual-level effects matter for the economic development of entire localities, we conduct our analysis at the city level. Further, we provide suggestive evidence on the mechanisms driving the effect.

To identify the causal effect of Internet availability on local economic growth, we exploit quasi-random variation in the timing of Internet availability induced by the arrival of the first sub-marine Internet cables (SMCs) in SSA in the early 2000s. This approach was established by Hjort and Poulsen (2019), who use an Internet speed upgrade in SSA induced by SMCs around 2010. In a difference-in-differences setting, we then compare the growth of cities with Internet access at the time of SMC arrival to a control group of similar cities getting access only somewhat later. Economic growth in each SSA city is measured by nighttime light intensity data captured by satellites, a well established proxy introduced by Henderson et al. (2011). City-level Internet availability is determined by data on the location of access points (APs) to the national Internet backbone.

[1] To date, SSA countries invested more than 28 billion US-Dollar into their national Internet backbone network (Hamilton Research, 2020). Facebook recently announced an effort to build a new sub-marine Internet cable to Africa for one billion US-Dollar (Bloomberg, 2000). And China plans to invest more than 60 billion US-Dollar in Africa's digital infrastructure as part of its Belt-and-Road initiative (Invesco, 2019).

We find that the connection to the Internet makes cities about 10% brighter in the years after connection, which approximately translates into 3 percentage points higher economic growth.[2] Moreover, we differentiate growth in the number of lit pixels, indicating a spatial expansion of cities, and growth in brightness, which is associated with a higher density of economic activity in the cities. We find that cities with Internet access are becoming brighter but not much larger. This provides suggestive evidence that cities with Internet access feature growing population density or increasing per capita value added, but do not grow much geographically.

We contribute to two main strands of the literature. First, we add to the literature assessing the impact of infrastructure investments on economic outcomes. There is ample evidence that investments in transportation infrastructure have long-lasting effects on regional growth and development (see e.g., Hornung, 2015; Allen and Arkolakis, 2019; Banerjee et al., 2020). The effect of digital infrastructure and especially (broadband) Internet has been assessed by Czernich et al. (2011) and Akerman et al. (2015), who show that firms and workers in connected regions become more productive when they have access to Internet. Closely related to our work is Hjort and Poulsen (2019), who study this in a developing-country context and find Internet to have a skill-biased and net positive employment effect in SSA. Our analysis contributes to these findings by showing that the benefits of digital infrastructure are present not only at the individual level but at the more aggregate city level as well. We further show that even the availability of very basic Internet speeds adopted by few individuals and businesses is beneficial for regional development.

Second, our work contributes to the literature on urban and regional development. Economic productivity is typically higher in cities for several reasons, for example, thick local labor markets, knowledge spillovers, low transportation costs, but also due to local provision of amenities (see e.g., Albouy, 2016; Clark et al., 2002; Deller et al., 2008). Thinking of digital infrastructure as local amenity, our findings indicate that the benefits of Internet availability are spatially highly concentrated, accruing to connected locations (cities) only, with important implications for regional inequality both across cities and between cities and rural areas.

The paper proceeds as follows. In Section 2, we provide a brief overview of the early Internet in Sub-Saharan Africa. Section 3 lays out the empirical strategy and in Section 4 the data and spatial methods are described. Results are presented in Section 5. Section 6 concludes.

2 Background

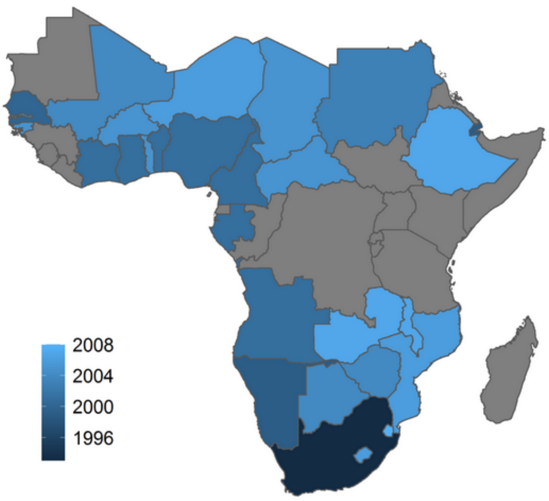
There are three major components of Internet infrastructure determining availability and bandwidth of Internet in a given location. First, international sub-marine fiber-optic cables connect SSA countries to the global Internet backbone. Second, within-country inter-regional fiber cables form the national backbone. Precondition for Internet availability in a location is an access point (AP) to the national backbone. Finally, individual users in a location are reached via the 'last mile' infrastructure.

[2] Henderson et al. (2011) finds an elasticity of GDP-to-light of 0.284.

2.1 International backbone: sub-marine cables

Since the vast majority of web pages and applications is hosted on servers located in North America or Europe, almost all African Internet traffic is routed inter-continentally (Kende and Rose, 2015; Chavula et al., 2015). Before the first fiber-optic sub-marine cables (SMCs) landed on African shores, the only way to connect to the Internet on the continent was via satellite.[3] While being largely unconstrained by geography and local infrastructure, satellite connection is very expensive and allows only for very narrow bandwidths. With SMCs—often a joint effort of governments, private investors, and/or multinational organizations—Internet connection was first brought to (Sub-Saharan) Africa at noticeable scale.

Figure 1: Sub-marine Cable Connection Years



As shown in Figure 1, the first wave of internet-enabled SMCs arrived in SSA countries only in the early 2000s.[4] These 'first-generation' cables had the capacity to provide Internet at basic speeds. The biggest of them was SAT-3 and started operating in 2001/2002. It featured landing points on the shores of eight SSA countries.[5] These landing points—typically one per country—constitute the starting point for the respective national backbones (cf. Section 2.2). Until the late 2000s, most SSA countries were connected to the Internet via SMC.

2.2 National backbone: inter-regional cables

After being routed through a SMC, Internet traffic travels through the national backbone. The national backbone infrastructure consists of inter-regional (fiber) cables. Many of these cables were built decades ago as part of the telegraph and telephone infrastructure and were only later used for the

[3] Single-channel and co-axial SMCs for telegraphy and telephony already existed before. The first telegraphy cable ('East coast' cable) started operating as early as 1879.

[4] South Africa was connected in 1993, when the first SMC constructed to enable the use of the Internet (SAT-2) preceded an old co-axial telephone cable established in 1968 (SAT-1).

[5] Benin, Cameroon, Côte d'Ivoire, Gabon, Ghana, Nigeria, Senegal, and South Africa.

transmission of early Internet traffic. They typically have been installed by the national telecom. Each (coastal) country typically has an own, self-contained backbone network.[6]

That is why, as soon as a new SMC arrives at a landing point of a SSA country, Internet becomes available country-wide in every location with access to the national backbone. As Internet capacity (i.e. speed) of the national backbone does not depend substantially on distance to the landing point, this upward shift occurs uniformly across the country's connected locations.

In the last decades, national backbone networks were continuously improved and expanded in parallel to the installation of SMCs. This network expansion focused heavily on connecting economically and/or politically important locations since they feature the largest market potential (high population density and GDP per capita).[7] This often lead to a network evolution where the national capital (often a coastal city) was connected first. Then, the network spread out to the next largest (or politically important) cities, which are often regional capitals or other large cities. We call all these cities 'nodal cities'.

Inter-regional cables are almost always built along pre-existing infrastructure (roads, but also railroads, pipelines, and electric cables) to minimize construction costs. Even though the goal was to connect nodal cities, cities on the route of inter-regional cables often got Internet access as well because establishing an 'access point' on an existing inter-regional cable entails little additional costs. That makes it often profitable to connect on-route cities between two nodal cities. Our empirical strategy (cf. Section 3) focuses on these on-route cities which get Internet connection because of their location next to an inter-regional cable but are not nodal cities themselves.

The timing of the network expansion via inter-regional cables is mainly determined by building costs and building speed. These factors are crucially impacted by topography and distance. Unfavorable topography (e.g. hard soil or rugged terrain) reduces building speed and makes it more expensive to install an additional kilometer of cable. Longer distance to a nodal city requires the installation of more kilometers of cable, which increases building speed and costs as well. Consequently, the exact connection year of on-route cities is determined mainly by these factors exogenous to decision makers.

2.3 Local transmission: 'last mile' infrastructure

Internet traffic transported by inter-regional cables is accessed at 'access points'. There are several technologies transmitting Internet traffic from these access points (so-called fiber nodes) to the user. These 'last mile' transmission technologies include fiber cables (FTTH/B), copper cables, and wireless transmission using cellular towers (e.g. mobile or WiMax).

Unlike in many developed nations which rely heavily on transmission to the end user via pre-existing telephony cable infrastructure, in SSA countries the Internet is mostly accessed via wireless and/or mobile devices. For this technology, no local cable network connecting each users exact locality (firm, household) is needed. Traffic data is exchanged wirelessly between cellular towers and the user's

[6] There are no network operators owning networks in more than one country. Cross-border connections are primarily established to connect landlocked neighboring countries.

[7] Routes establishing connections to (landlocked) neighboring countries are a focus of network expansion as well.

device. Relative to the costs to construct an inter-regional cable it is thus cheap to establish Internet access along the cable, making it profitable for the network operator to establish access points even in on-route cities, which are typically much smaller than nodal cities.

3 Empirical strategy

We are interested in the relationship between Internet availability and local economic growth. However, this correlation is not informative about the causal effect of Internet availability on local economic growth due to endogeneity concerns. In particular, cities with and without Internet access might be very different. Internet access is not randomly assigned and likely driven by commercial interest and/or political and administrative planning. The connection of a priori larger cities on a higher growth path may be prioritized above smaller cities which are stagnating.

To address these endogeneity concerns, we leverage two distinct features of Internet infrastructure evolution in SSA countries. First, we use plausibly exogenous time variation in connections to submarine cables, which determine Internet availability nation-wide, to investigate the effect of Internet availability. Second, we focus on on-route cities between two nodal cities that are getting connected eventually.

We focus our analysis on on-route cities eventually getting an access point for two reasons. First, they are very similar to each other in key characteristics such as size and infrastructure. They all get connected eventually, mainly because of their favourable location between two nodal cities. Second, the exact timing of the connection of on-route cities cannot be influenced by decision makers because the construction speed is determined by topographical and geographical factors (cf. Section 2.2).

Furthermore, we follow Hjort and Poulsen (2019) and argue that the exact timing of SMC arrival is essentially random. First, for each individual SMC, it is exogenous to national planners when exactly it is put into operation and starts providing (international) Internet connection. This is because each SMC typically connects many countries and the connection years are highly uncertain both due to unforeseen delays in construction and coordination difficulties among consortium members.[8] Second, across different SMCs the connection years are mainly determined by the geographical location because SMCs come from Europe and connect SSA countries according to their location at Africa's coastline.

This enables us to use a difference-in-differences (DiD) design. We compare on-route cities that already have access to the national backbone when the SMC arrives to on-route cities getting an access point in later years (first difference) before and after the arrival of a SMC (second difference). Since national backbones are self-contained and (coastal) countries have own landing points (cf. Section 2.2) each country has a specific treatment date.[9] Therefore—depending on the nation-wide connection date—on-route cities are treated at different points in time.

[8] For example, the cable EASSy was delayed by five years due to coordination difficulties among consortium members (**?**).

[9] Landlocked countries connected via a neighboring coastal country feature the same treatment date as the country from which it gets the connection.

The basic model used to identify the average treatment effect on the treated (ATT) of Internet availability on local economic growth is given by

$$y_{ict} = \beta_0 + \beta_1 \left(smc_{ct} \times access_{ic} \right) + \boldsymbol{X}_{ict}\beta + \alpha_{ic} + \delta_{ct} + \epsilon_{ict} \tag{3.1}$$

where y_{ict} is economic growth of city i in country c in calendar year t as proxied by nighttime-light luminosity measured by satellites (cf. Section 4). The dummy variable smc_{ct} indicates if country c has access to a SMC in calendar year t. The variable $access_{ic}$ is one if city i in country c has an access point that was established in the year of SMC arrival or before. Cities getting a node in the four years under observation after the arrival of the SMC are excluded from the control group. Thus, the interaction term $smc_{ct} \times access_{ic}$ indicates if a city has had an access point in the year of SMC arrival or if it got an access point at least five years later. $\boldsymbol{X}_{ict}$ contains time-varying control variables.

We include two types of fixed effects into the model. Time-constant differences across cities are captured by city fixed effects α_{ic}. Differences across calendar years common to all cities within a country are absorbed by country-year fixed effects δ_{ct}. Note that this allows for country-specific time trends and variations in satellite sensor quality over years. ϵ_{ict} is an error term.

The coefficient of interest is β_1. It captures the effect of Internet availability on local economic activity. Like in many other DiD applications, our (panel) data are serially correlated in the time dimension. Hence, we use cluster-robust standard errors whereby we cluster at the city level.

The key identifying assumption for this DiD model is that treatment and control cities would have evolved similarly in the absence of treatment (parallel trends assumption). This assumption cannot be tested. Its plausibility can, however, be examined by testing for pre-treatment differences in time trends between treatment and control group. Therefore, we look at the dynamic impact of Internet availability on local economic activity using event studies:

$$y_{ict} = \beta_0 + \sum_{j=\underline{T}}^{\overline{T}} \beta_{1j} \left(t_j \times access_{ic} \right) + \boldsymbol{X}_{ict}\beta + \alpha_{ic} + \delta_{ct} + \epsilon_{ict} \tag{3.2}$$

where t_j indicates the year relative to treatment year—i.e. the year when the (new) SMC arrives—starting in relative year $j = \underline{T}$ and ending with relative year $j = \overline{T}$. Treatment year is normalized to $j = 0$. We exclude $j = -1$ as the reference point. Thus, the interaction $t_j \times node_{ic}$ indicates if city i in country c is part of the treatment group and restricts the coefficient to one particular relative year j.

The coefficients β_{1j} inform about the dynamic effect of broadband availability. Thereby, each coefficient captures relative-year-specific treatment effects. We expect to see no effect before the treatment. Thus, if we cannot distinguish the estimates of the coefficients on the pre-treatment relative-year dummies (i.e. $\hat{\beta}_{1j} \forall j < 0$) from zero, treatment and control group follow similar trends before the treatment, supporting the common trends assumption.

4 Data and spatial methods

We analyze the effect of Internet availability on local economic growth in SSA.[10] To this end, we tap two main data sources. First, local economic activity is measured by nighttime-light (NTL) satellite data. Second, locations connected to the Internet are identified via the geo-location of access points (APs). This section describes the data sets we use, the preprocessing steps and spatial methods applied, and shows descriptive statistics relevant for our analysis.

4.1 Local economic activity: nighttime-lights

Since geographically and chronologically granular data on economic activity in SSA is lacking, we deploy nighttime-light (NTL) data. This data makes it possible to measure human-caused nighttime-light emission in a geographically high resolution and on a yearly basis. There have been two major programs that collected NTL data. First NTL data was collected in the *Defense Meteorological Satellite Program* (DMSP) *Operational Linescan System* (OLS) between 1992 and 2013. The follow-up program *Visible Infrared Imaging Radiometer Suite* (VIIRS) on the *Suomi National Polar-orbiting Partnership* satellite started in April 2012. We use the harmonization of the two sources by Li et al. (2020) to get consistent yearly NTL data from 1992 to 2018.

The original purpose of the DMPS-OLS was meteorological observation. Hence, the NTL for measuring human activity is only a byproduct. The instruments of DMPS-OLS satellites measure light intensity on an integer scale from 0 to 63 with pixels covering 30 arc-second grid-cells (an area of .86 km^2 at the equator). The data is then combined to yearly composite images. With VIIRS this changed. Now, both the spatial (pixels are smaller) and radiometric resolutions (both dark and bright spots are recorded better) have been improved. Additionally, the time resolution improved as well (monthly).

To harmonize these two sources, Li et al. (2020) recalculated the newer and better VIIRS data to the (spatial and radiometric) resolution of the original DMPS-OLS data. They exploit the overlay of the two programs in 2013. In this procedure, noise from aurora, fires, boats, and other temporal lights was also excluded. Another advantage of this data is that the DMPS-OLS data has been inter-calibrated globally from 1992 to 2013 as well, making it temporally consistent. Unlike in the developed world, very high light intensities (i.e. top-coded pixels) are less a concern in the context of SSA.[11] Therefore, the lower radiometric resolution is applicable in our case.

On the country level, NTL data is well established as a measure of economic activity and widely used by economists (Henderson et al. (2012) and Chen and Nordhaus (2011) among the first ones). Closely related to our work, Storeygard (2016) established this data on city-level. On a broader area, Bruederle and Hodler (2018) added the relation to household wealth, education, and health for DHS cluster locations as well as for grid cells of roughly 50×50 km.

[10] We exclude all islands except for Madagascar: Mayotte, Réunion, St. Helena, Madeira, the Canary Islands, Cape Verde, Comoros, Mauritius, the Seychelles, and Sao Tome and Principe. Finally, we do not analyze South Africa due to data unavailability as it was connected very early in 1993.

[11] Looking at the sample that only contains cities, usually less than 2% of the pixels are assigned to at least 60.

Within each city, we define several outcome measures. Local economic activity is measured by summing the light intensity of all pixels within a city in each year. This measure was already applied by Storeygard (2016) and accounts for both geographical extension and light intensity. As alternative measures, we calculate the sum of all lit pixels, instead of their intensity, in each year within a city and the average light intensity of pixels within each city and year. We interpret the sum of lit pixels as a proxy for spatial extension of a city and the average light intensity as a proxy for density in terms of population or per capita economic activity. Since low light intensities are prone to measurement error of the satellite's sensors we calculate these measures applying thresholds on the light intensity (minimum luminosity of 10). This way, we can also learn more about the mechanism. [12]

We apply the NTL data also to define cities. To measure each city's evolution over time, we define a city as a cluster of contiguous lit pixels and track them over time. We overlay NTLs from all years and define contiguous lit pixels as a city.[13] That means, a city is defined by the maximum spatial expansion of a light cluster.[14] This approach is similar to Storeygard (2016) and Henderson et al. (2017).

For a robustness check, we take established data from *Africapolis*.[15] This database contains the geographical delineation of 5,811 SSA cities with more than 10,000 inhabitants in 2015. The delineation is based on a classification algorithm detecting built-up areas. The median city size is around 20.000 inhabitants and about 90 percent of all cities have less than 100.000 inhabitants.

4.2 Backbone infrastructure: access points and sub-marine cables

The geo-location of the access points (APs) in the national backbone comes from *Africa Bandwidth Maps*.[16] The database contains the most comprehensive set of APs for Africa and covers the period from 2009 to 2019. *Africa Bandwidth Maps* directly sources its data from the network operators.

As APs existing in 2009 were mostly established earlier, we conduct an extensive review of network deployment projects for each country. Thereby, we are able to determine the construction years of the nodes from 2009 going back to the late 1990s.This makes it possible for us to identify locations with Internet access, when the first SMCs arrived in the first decade of the 21th century. By matching the APs to the cities via their geo-location, we are able to identify the broadband Internet access year for each city.

Figure A.1 maps all 2.708 APs in the countries we analyze and their respective year of construction. About half of them were constructed since 2013. Especially in bigger cities, more than one AP is usually built to account for the limited capacities of each AP. Around 900 APs were built in different cities. In 2019, although 189 new APs were constructed, only 27 new cities were exploited. Before 2009, 87 (regional) capitals were connected, that is about one third of the connected cities. Additionally, 26 cities with more than 100.000 inhabitants were connected before 2009. These cities fall into our definition of 'nodal cities' (c.f. Section 3) and are not considered in the analysis.

[12] As specified in Section 3, we apply the logarithm of each outcome measure.

[13] For the analysis of the 'first-generation' Internet introduction, we take the years between 1995 and 2008 as observation period. Because often suburb-like lit areas are found just a few pixels apart from another contiguously lit area, we allow three unlit pixels between lit pixels.

[14] This makes the underlying area of a city time invariant.

[15] https://africapolis.org

[16] http://www.africabandwidthmaps.com

To measure time of treatment, we use information on SMCs landing on the shores of SSA countries.[17] In particular, we use the country-specific connection dates and the geo-location of the landing points.[18]

When building the national backbone networks, the goal was to connect major cities and (regional) capitals ('nodal cities'; cf. Section 2.2). To identify to which cities, towns, or (large) villages the nighttime-light clusters correspond, we use data on populated places from the *Natural Earth Database*. Matching these data to the cities enables us to flag which cities are (regional) capitals and/or satisfy a minimum population size (100,000 inhabitants). We define these cities as 'nodal cities'. We derive information on economic relevance, by population size, and political relevance, for (regional) capitals.

4.3 Descriptive statistics

Before coming to the estimation results, we provide descriptive statistics to compare treated and untreated cities. Table A.1 shows the connection years of the different generations of SMCs by country. We focus on early SMCs, and therefore, drop countries which were connected after 2008 for the first time.[19] This leaves us with 27 countries. Among the first countries are Djibouti, where an SMC landed in 1999, Namibia, which was connected by a trans-national fiber cable from South Africa also in 1999, and Senegal, which was connect in 2000 with an individual-country SMC. In 2001, seven more countries were connected by a single SMC, the SAT-3 cable. In the following years, 17 more countries got an SMC connection or were connected through their neighboring countries until 2008.

However, not all countries that were connected until 2008 had built a national backbone infrastructure when the respective SMC or the connection through a neighboring country arrived. In this case, the treatment group is missing as in no city Internet was available right after the connection. This reduces the number of countries to 24 in our analysis.[20] Moreover, there are 11 countries which built at least one AP before the SMC arrival but only in 'nodal cities'.[21] Finally, Namibia has to be dropped as it did not construct APs after getting Internet access. Here, we cannot define a control group. This leaves us still with 12 countries, on which we will estimate.

This sample is not necessarily balanced. From the sample, the first connected country is Senegal, which was connected in 2000, which leaves us with 7 pre-treatment years. Hence, on the left with a truncation at -7 years, we have no data loss on the country-side. Though, it limits our pre-treatment comparison to a period of seven years. Malawi, Mozambique only have 2, and Zambia only 3, post-treatment years. Hence, estimating on a balanced sample with either 3 or 4 post-treatment years, leaves us with a sample of 10 or 9 countries. We present the following descriptive statistics on the

[17] The data comes primarily from *Submarine Cable Map* (https://www.submarinecablemap.com).

[18] A broader overview is given in Section 2.1

[19] The next generation SMCs landed between 2009 and 2012 (c.f. Section 2.1).

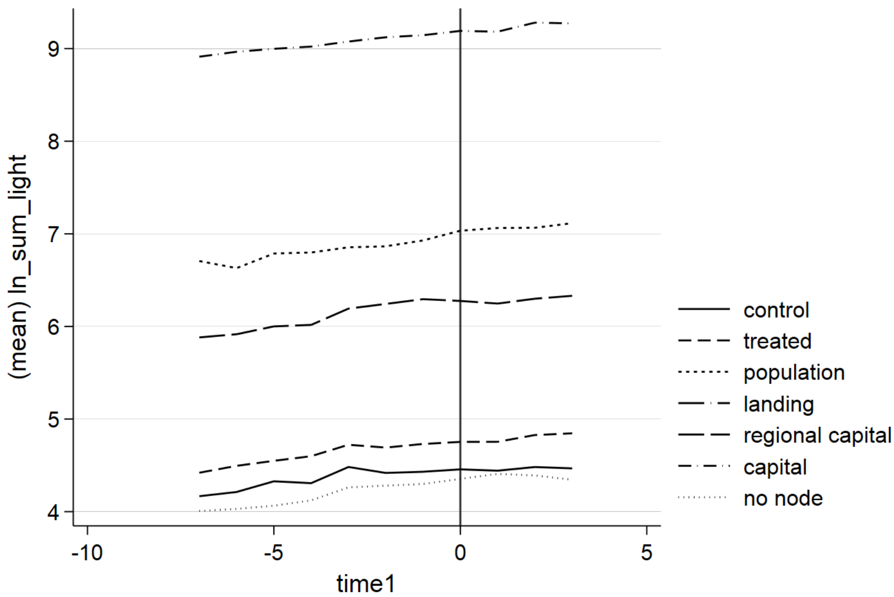
[20] Central African Republic has not yet built a national backbone infrastructure. In Djibouti, the first APs were built in 2007, which is 8 years after the first SMC connection. Nigeria built its first APs in 2003, which is 2 years after the arrival of the first SMC. This means especially that the biggest two SSA economies, South Africa and Nigeria, are not included in our analysis.

[21] These countries are: Angola, Chad, Côte d'Ivoire, Cameroon, Gabon, Gambia, Ghana, Niger, Guinea-Bissau, Lesotho, and Swaziland. The last three built all APs until today only in 'nodal cities'. Guinea-Bissau, for instance, only built APs in 2005, the connection year.

balanced sample of 10 countries between time -7 and 3. Other (un)balanced samples are used for robustness in Section 5.[22]

We start the descriptive statistics with a raw data plot of the whole sample (Figure 2) and then focus on the treatment and control group (Figure 3). In Figure 2, besides the treatment and control group, all from our analysis the excluded cities are shown. It can be seen that the capitals are by far the biggest citites and that cities that do not get an access point by the end of our data are the smallest on average. However, they are not a lot smaller than the cities in the control group. Additionally, the other excluded cities, the regional capitals and the cities with an population of at least 100,000 inhabitants are rather similar in comparison to the cities in our analysis. It can be noted that the regional capitals on average do not exceed the bigger cities category. It is hard to see growth patterns in this aggregated figure. Therefore, we show a figure of just the treatment and control group next. Figure 3 shows that in the early pre-treatment years both groups grew with similar rates. In more recent years before the treatment, this growth stagnated. This stagnation holds on for the control group, while the treated cities grew again in the second and third year after the treatment.

Figure 2: Time Trends by Sample



Next, we compare both groups in their characteristics (Table **??**). Variables are either geographic, whether they are located at a national border, at the coast, or at a river, or they compare the local infrastructure, whether there is an airport, a port, a bank, a specific type of (rail)road, or educational or health institutions. We can show that both groups are very similar. The biggest difference is that in the treatment group, more cities are located at the coast than in the control group. The variables come either from *Natural Earth* or *Open Street Map*. Therefore, the data is time-invariant and not dated before the treatment but was collected in 2010 or more recently.

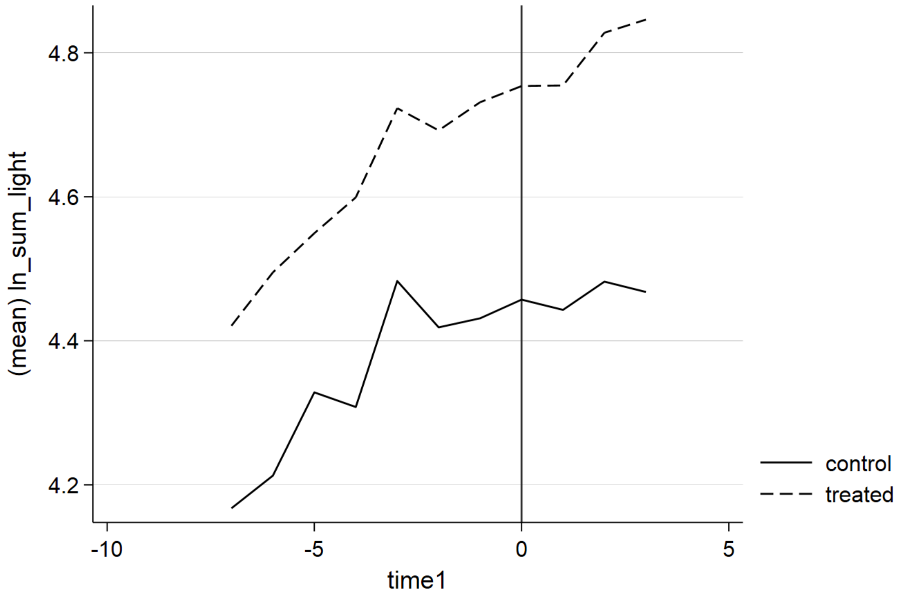
[22] Descriptive statistics on these samples are available upon request.

Table 1: Means Comparison between Treatment and Control Group

	Mean		
	Control	Treated	Δ
Area	9.910	10.832	0.923
Population			
2015	39,766	38,316	-1,450
2010	30,140	29,742	-397
2000	20,918	19,569	-1,349
Health			
Hospital	0.162	0.115	-0.047
Pharmacy	0.206	0.246	0.040
Doctor	0.029	0.131	0.102**
Transportation			
Railroads_ne	0.294	0.410	0.116
Roads (natural earth)	0.721	0.738	0.017
Roads (major 1)	0.000	0.033	0.033
Roads (major 2)	0.456	0.492	0.036
Border crossings	0.000	0.016	0.016
Coast	0.044	0.180	0.136**
River	0.059	0.049	-0.010
Port	0.000	0.000	0.000
Education & Other			
School	0.324	0.410	0.086
College	0.088	0.115	0.027
University	0.015	0.049	0.034
Bank	0.382	0.410	0.027
Cities	68	61	

Notes: Asterisks in column Δ report t-test for equality of means: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 3: Time Trends of Treatment and Control Group



We focus our analysis on mid-sized cities. Therefore, we set a focus on the comparison of the population between the two groups. The population data is coming with the built-up areas from *Africapolis*. Figure 5 shows, first, that both groups are very similar in their population distribution, and second, that mid-sized cities have a population of around 30,000 inhabitants. There are 285 cities excluded from the analysis, either because they are 'nodal cities' or cities that did not build an AP until 2019. These cities are on average (144.000 inhabitants) bigger than the cities of our analysis. 67 cities of them have more than 100.000 inhabitants.

Finally, we show the geographic distribution of the treatment and control group cities. For many countries, there is no specific pattern. Sudan is a special case as 10 cities are in the control group but only one city is in the treatment group. One should also note that of the 10 countries 4 are coastal, while 6 are landlocked.

Figure 4: Population Density in Treated and Control Cities

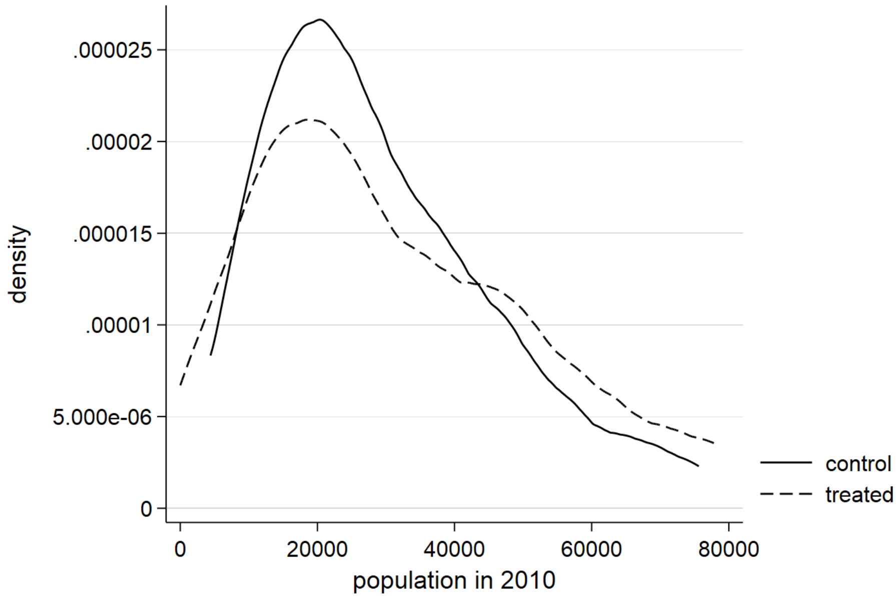
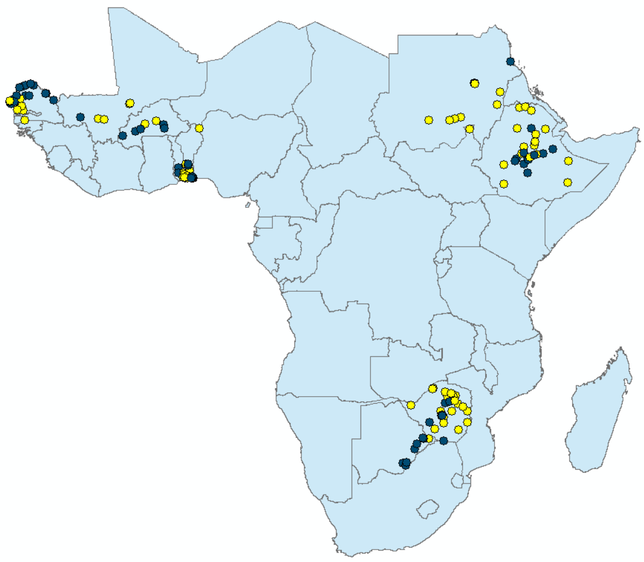


Figure 5: City Locations in the Estimation Sample



Notes: Blue dots are treated, yellow dots are in the control group.

5 Results

5.1 Main effects

We estimate the effect of Internet availability on local economic growth. We are particularly interested in the effect of early Internet availability brought by the 'first-generation' SMCs. First, we estimate a difference-in-differences approach on a balanced panel. Depending on the country only very few years lie between the two generations.[23] We include only countries with at least 4 years after the 'first' and before a 'second-generation' SMC arrives.

In our main specification, we take the logarithm of the sum of luminosity as measurement. Table 2 shows the main results. In the first column, the raw specification, nodal cities and the landing point are still included. We then eliminate step-wise these nodal cities until we reach our preferred specification where the cities are very similar and without an obvious reason why especially these cities get an access point. In the second column, we remove the city of the landing point and the national capital. In the third column, we also remove regional capitals. In the fourth column, we remove cities of more than 100.000 inhabitants. Finally, we add as a control variable the city's relative mobile coverage.

We find a highly significant positive effect. In our preferred specification, cities which were connected to the Internet are 10% brighter than cities without Internet availability. In economic terms, this corresponds to a 3 percentage point higher GDP growth. For this back of the envelope calculation, we take the GDP light intensity elasticity of $\epsilon_{GPD,light} = 0.284$ from Henderson et al. (2012).

Table 2: The Effect of Internet on Economic Growth of Cities

Dep. var.: light intensity	(1)	(2)	(3)	(4)	(5)
post × treated	0.0595**	0.0571*	0.0845**	0.100**	0.102***
	(0.0289)	(0.0296)	(0.0346)	(0.0385)	(0.0386)
GSM coverage					0.0183
					(0.0384)
City FE	✓	✓	✓	✓	✓
Country × year FE	✓	✓	✓	✓	✓
City restriction					
Landing point and capital		✓	✓	✓	✓
Regional capitals			✓	✓	✓
Population >100k				✓	✓
Observations	2,244	2,136	1,788	1,548	1,548
#Treated	108	99	78	61	61
#Control	79	79	71	68	68
Adj. R-squared	0.990	0.986	0.984	0.974	0.974

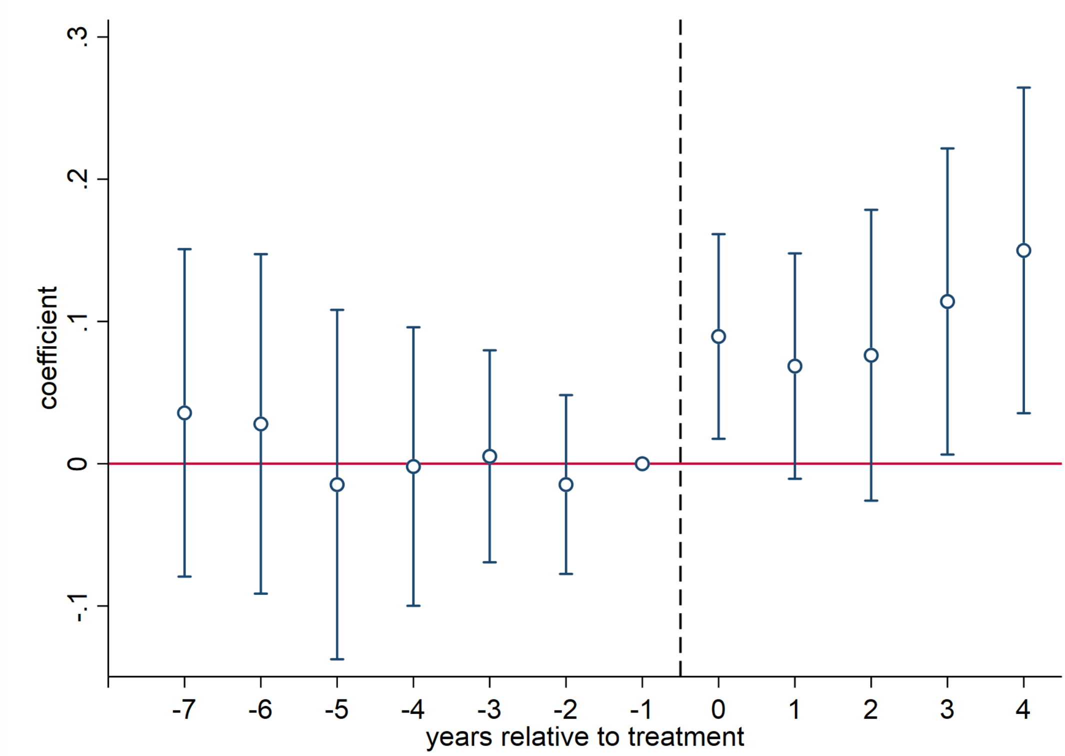
Notes: Light intensity is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within city area. GSM mobile phone coverage is calculated as the percentage share of city area covered with signal. Sample successively restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100k inhabitants. Robust standard errors clustered by city reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For our preferred specification, Figure 6 presents event study coefficients. Before the SMC connection, the estimates are close to zero and insignificant. The effect kicks in directly after the SMC landed. This is not surprising as the national backbone infrastructure is already existing when the SMC

[23] For instance, Mozambique was first connected in 2006 and upgraded in 2009. Eswatini, even, was connected first in 2008 and got an upgrade only one year later.

arrives. In the following years, the estimates are constantly around .1 and have slightly the tendency to increase to .15 in year four after the connection.

Figure 6: Event Study Coefficients



Notes: Coefficients for event study specification of model (5) from Table 2. Robust standard errors clustered by city for 95% confidence interval reported as bars.

5.2 Mechanism

As an alternative outcome measure to the sum of luminosity (Table 3, column 1 & 2), we also take the mean luminosity (Table 3, column 3 & 4) and the sum of lit pixels (Table 3, column 5 & 6), its 'size'. All outcomes are logged. We also present estimation results restricted to brighter pixels, pixels with an intensity of more than 10. We observe cities with Internet to become brighter (column 3) and also the number of bright pixels increases (column 6). Generally, the effect for the brighter pixels is stronger than the effect for all pixels. As brightness is a proxy for density, we assume that growth is triggered through activity in the already more developed parts of the city. This makes totally sense, as we expect here the Internet to have the strongest impact. New lit pixels might be relatively dark, and therefore, they might lower the mean. Each new lit pixel increases the sum of lit pixels by one. As light blurs on the satellite images, lit points at the border of the satellite image are out of border in the built-up area. Therefore, 'size' measures the growth towards the restrictions of the city. Hence, the growth can mainly be reduced to an increase in light intensity.

Table 3: Population, Intensive, and Extensive Margin

	growth proxy		intensive		extensive	
	(1)	(2)	(3)	(4)	(5)	(6)
post × treated	0.102***	0.0941**	0.0935***	0.0871**	0.0272**	0.0238*
	(0.0386)	(0.0391)	(0.0348)	(0.0354)	(0.0134)	(0.0126)
GSM coverage	0.0183	0.0136	0.0198	0.0157	-0.0101	-0.0123
	(0.0384)	(0.0388)	(0.0341)	(0.0345)	(0.0152)	(0.0153)
Population		0.504*		0.436*		0.232*
		(0.294)		(0.254)		(0.129)
City FE	✓	✓	✓	✓	✓	✓
Country × year FE	✓	✓	✓	✓	✓	✓
City restriction	✓	✓	✓	✓	✓	✓
Observations	1,548	1,548	1,548	1,548	1,548	1,548
#Treated	61	61	61	61	61	61
#Control	68	68	68	68	68	68
Adj. R-squared	0.974	0.974	0.918	0.919	0.985	0.985

Notes: Light intensity in the combined models is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within city area. The measure for the intensive margin is logarithmic mean light intensity and for the extensive margin logarithmic sum of pixels. GSM mobile phone coverage is calculated as the percentage share of city area covered with signal. Population is measured as inhabitants per square kilometer within city area from GPW. Sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100k inhabitants. Robust standard errors clustered by city reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We plan to investigate further mechanisms. Besides the different outcomes, we plan to investigate spatial inequality. Here, we want to analyze whether the hinterland gets darker where cities grow due to Internet availability.

5.3 Heterogeneity and Robustness

First, we analyze different connection years. In Table 4, we first remove later connection years. Here, the estimate increases slightly. Moreover, investigating only countries that where connected in 2001, the year where several countries were connected by one bigger SMC, shows a strong positive effect of Internet availability on local economic growth. Also, when dropping countries with an early connection year step-wise the effect size remains but drops slightly in the level of significance.

Table 4: Heterogeneity by SMC Connection Year

Dep. var.: light intensity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
post × treated	0.100**	0.142***	0.129**	0.115**	0.123**	0.196**	0.110**	0.0876*	0.0876*	0.0907*
	(0.0385)	(0.0466)	(0.0509)	(0.0548)	(0.0591)	(0.0776)	(0.0430)	(0.0503)	(0.0503)	(0.0529)
City FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country × year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
City restriction	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Connected before	2008	2006	2005	2004	2003	–	–	–	–	–
Connected in	–	–	–	–	–	2001	2001	2001	2001	2001
Connected after	–	–	–	–	–	–	2000	2001	2002	2003
Observations	1,548	1,212	1,092	636	504	216	1,260	1,044	1,044	912
Adj. R-squared	0.974	0.977	0.978	0.962	0.963	0.960	0.972	0.972	0.972	0.976

Notes: Light intensity in the combined models is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within city area. GSM mobile phone coverage is calculated as the percentage share of city area covered with signal. Population is measured as inhabitants per square kilometer within city area from GPW. Sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100k inhabitants. Robust standard errors clustered by city reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We define the treatment for a city as having an AP within 10 km. In Table 5, we vary this distance. The estimate remains around .1.

Table 5: Robustness: Distance to Access Point

Dep. var.: light intensity	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
post × treated	0.100**	0.0766	0.116**	0.113**	0.110**	0.0922**	0.112***	0.111***	0.0876**	0.102***	0.101***
	(0.0385)	(0.0606)	(0.0462)	(0.0435)	(0.0430)	(0.0403)	(0.0361)	(0.0359)	(0.0376)	(0.0354)	(0.0359)
Distance to AP	10 km	0 km	2 km	4 km	6 km	8 km	12 km	14 km	16 km	18 km	20 km
City FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country x Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
City restriction	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,548	768	1,128	1,320	1,404	1,488	1,620	1,644	1,572	1,620	1,644
Adj. R-squared	0.974	0.976	0.976	0.973	0.973	0.974	0.973	0.974	0.974	0.974	0.974

Notes: Light intensity in the combined models is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within city area. GSM mobile phone coverage is calculated as the percentage share of city area covered with signal. Population is measured as inhabitants per square kilometer within city area from GPW. Sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100k inhabitants. Robust standard errors clustered by city reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Robustness: Panel Balancing

Dep. var.: light intensity	(1)	(2)	(3)	(4)	(5)
post × treated	0.0941**	0.101***	0.0750**	0.0765**	0.0675**
	(0.0391)	(0.0380)	(0.0304)	(0.0297)	(0.0282)
GSM coverage	0.0136	0.0263	0.0280	0.0383	0.0353
	(0.0388)	(0.0407)	(0.0291)	(0.0300)	(0.0270)
Population	0.504*	0.463*	0.422**	0.388**	0.475***
	(0.294)	(0.274)	(0.180)	(0.165)	(0.153)
Time frame	-7 to 4	-6 to 4	-7 to 3	-6 to 3	-6 to 3
Balanced panel	✓	✓	✓	✓	
City FE	✓	✓	✓	✓	✓
Country x Year FE	✓	✓	✓	✓	✓
City restriction	✓	✓	✓	✓	✓
Observations	1,548	1,419	1,804	1,640	1,910
R-squared	0.974	0.975	0.976	0.977	0.975

Notes: Light intensity in the combined models is measured as the logarithmic sum of light intensities of DMPS-OLS pixels within city area. GSM mobile phone coverage is calculated as the percentage share of city area covered with signal. Population is measured as inhabitants per square kilometer within city area from GPW. Sample restricted by excluding landing point cities and capitals, regional capitals, and cities with more than 100k inhabitants. Robust standard errors clustered by city reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We plan further robustness checks, where we perform regressions on a different definition of the city. We show the same measures as in Table **??**. Instead of built-up areas, we take the maximal expansion of lit pixels over time as the agglomeration area as described in Section 4. Therefore, the estimate of the outcome ‘size’ is more meaningful.

Finally, we plan to split the sample into coastal and landlocked countries. Coastal countries get connected by SMCs, while landlocked countries could only receive a connection through other (coastal) countries. Therefore, landlocked countries depend on their neighboring countries and cannot influence the year they receive an Internet connection.

6 Conclusion

We find that the connection to the world wide web makes cities about 10 % brighter. That is comparable to a growth of 3 percentage points in terms of GDP. Moreover, we can differentiate growth in more pixels, where cities would increase in their area, and in a higher average of the light intensity, which is generally associated with a higher density of the cities. We find that cities with Internet availability are becoming brighter on average and get more bright pixels. We conclude, that the GDP growth is associated with city becoming denser, probably in the city centers. This is exactly where we would expect Internet availability to have the strongest effect.

References

Akerman, A., I. Gaarder, and M. Mogstad (2015, nov). The skill complementarity of broadband internet. *Quarterly Journal of Economics 130*(4), 1781–1824.

Albouy, D. (2016). What are cities worth? land rents, local productivity, and the total value of amenities. *The Review of Economics and Statistics 98*(3), 477–487.

Allen, T. and C. Arkolakis (2019). The Welfare Effects of Transportation Infrastructure Improvements. *National Bureau of Economic Research*.

Banerjee, A., E. Duflo, and N. Qian (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics 145*(January), 102442.

Bloomberg (2000). Faster internet coming to africa with facebook's $1 billion cable.

Bruederle, A. and R. Hodler (2018). Nighttime lights as a proxy for human development at the local level. *PLoS ONE 13*(9), 1–22.

Chavula, J., N. Feamster, A. Bagula, and H. Suleman (2015). Quantifying the effects of circuitous routes on the latency of intra-africa internet traffic: A study of research and education networks. In A. Nungu, B. Pehrson, and J. Sansa-Otim (Eds.), *e-Infrastructure and e-Services for Developing Countries*, Cham, pp. 64–73. Springer International Publishing.

Chen, X. and W. D. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences of the United States of America 108*(21), 8589–8594.

Clark, T. N., R. Lloyd, K. K. Wong, and P. Jain (2002). Amenities Drive Urban Growth. *Journal of Urban Affairs 24*(5), 493–515.

Czernich, N., O. Falck, T. Kretschmer, and L. Woessmann (2011). Broadband Infrastructure and Economic Growth. *Economic Journal 121*(552), 505–532.

Deller, S. C., V. Lledo, and D. W. Marcouiller (2008). Modeling regional economic growth with a focus on amenities. *Review of Urban and Regional Development Studies 20*(1), 1–21.

Hamilton Research (2020). Africa bandwidth maps.

Henderson, J. V., A. Storeygard, and U. Deichmann (2017). Has climate change driven urbanization in Africa? *Journal of Development Economics 124*, 60–82.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Henderson, V., A. Storeygard, and D. N. Weil (2011). A bright idea for measuring economic growth. *American Economic Review 101*(3), 194–199.

Hjort, J. and J. Poulsen (2019). The arrival of fast internet and employment in Africa. *American Economic Review 109*(3), 1032–1079.

Hornung, E. (2015). Railroads and growth in Prussia. *Journal of the European Economic Association 13*(4), 699–736.

Invesco (2019). China embarks on digital silk road.

Kende, M. and K. Rose (2015). Promoting Local Content Hosting to Develop the Internet Ecosystem. Technical report, Internet Society.

Li, X., Y. Zhou, M. Zhao, and X. Zhao (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific Data 7*(1), 1–9.

Storeygard, A. (2016). Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa. *Review of Economic Studies 83*(3), 1263–1295.

World Bank (2016). World Development Report 2016: Digital Dividends. Technical report.

A Appendix

Table A.1: Overview of SMC arrival years

country	countrycode	cntry	location	access_year_initial	access_year_submarine	comment	upgrade_year_submarine
Angola	AO	ao	coast	2001	2001		2012
Benin	BJ	bj	coast	2001	2001		2012
Botswana	BW	bw	landlocked	2004		South Africa	2009
Burkina Faso	BF	bf	landlocked	2005		Senegal-Mali	2010
Burundi	BI	bi	landlocked	2012			2012
Cameroon	CM	cm	coast	2001	2001		2012
Central African Republic	CF	cf	landlocked	2005			2012
Chad	TD	td	landlocked	2005		Cameroon	2012
Congo	CG	cg	coast				2012
Côte d'Ivoire	CI	ci	coast	2001	2001		2010
Democratic Republic of the Congo	CD	cd	coast	2012	2012		2012
Djibouti	DJ	dj	coast	1999	1999		2009
Equatorial Guinea	GQ	gq	coast	2012	2012		2012
Eritrea	ER	er	coast	2009	2009		2009
Ethiopia	ET	et	landlocked	2007		Sudan	2012
Gabon	GA	ga	coast	2001	2001		2012
Gambia	GM	gm	coast	2005	2012	Senegal	2012
Ghana	GH	gh	coast	2001	2001		2010
Guinea	GN	gn	coast	2012	2012		2012
Guinea-Bissau	GW	gw	coast	2005	2012	Senegal	2012
Kenya	KE	ke	coast	2009	2009		2009
Lesotho	LS	ls	landlocked	2006		South Africa	2010
Liberia	LR	lr	coast	2012	2012		2012
Madagascar	MG	mg	island	2009	2009	La Reunion	2009
Malawi	MW	mw	landlocked	2007		Mozambique	2010
Mali	ML	ml	landlocked	2004		Senegal	2010
Mozambique	MZ	mz	coast	2006	2009	South Africa	2009
Namibia	NA	na	coast	1999	2012	South Africa	2012
Niger	NE	ne	landlocked	2006		Senegal-Mali-Burkina-Faso	2012
Nigeria	NG	ng	coast	2001	2001		2010
Réunion	RE	re	island	2002	2002		2009
Rwanda	RW	rw	landlocked	2009		Kenya-Uganda	2010
Senegal	SN	sn	coast	2000	2000		2010
Sierra Leone	SL	sl	coast	2012	2012		2012
Somalia	SO	so	coast	2010	2010		2010
South Africa	ZA	za	coast	1993	1993		2009
Sudan	SD	sd	coast	2003	2003		2010
Swaziland	SZ	sz	landlocked	2008		South Africa	2009
Tanzania	TZ	tz	coast	2009	2009		2009
Togo	TG	tg	coast	2005		Burkina Faso	2012
Uganda	UG	ug	landlocked	2009		Kenya	2009
Zambia	ZM	zm	landlocked	2007		Zimbabwe	2011
Zimbabwe	ZW	zw	landlocked	2004		South Africa	2011
Mauritania	MR	mr	coast				2012

Figure A.1: Fiber Nodes and Their Year of Construction