



HAL
open science

Broadband infrastructure deployment, digital vulnerability, and local firm performance in developing and transition countries

Joel Cariolle, Maëlan Le Goff, Olivier Santoni

► To cite this version:

Joel Cariolle, Maëlan Le Goff, Olivier Santoni. Broadband infrastructure deployment, digital vulnerability, and local firm performance in developing and transition countries. 2018. <hal-01758660v2>

HAL Id: hal-01758660

<https://hal.science/hal-01758660v2>

Submitted on 24 Apr 2018

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


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



HAL Authorization

Broadband infrastructure deployment, digital vulnerability, and local firm performance in developing and transition countries

Joël CARIOLLE, Maëlan LE GOFF, and Olivier SANTONI

 JOËL CARIOLLE, Research officer at FERDI.
E-mail : joel.cariolle@ferdi.fr (corresponding author)

 MAËLAN LE GOFF, Research officer at Banque de France.
E-mail : maelan.legoff@banque-france.fr

 OLIVIER SANTONI is Geomatician at FERDI.
E-mail : olivier.santoni@ferdi.fr

Abstract

This paper provides evidence on the impact of fast Internet on firm performance in developing and transition economies. Over the last three decades, international connectivity has been boosted by the laying of more than 300 submarine telecommunications cables (SMC). Almost all coastal developing and transition countries are plugged into the global Internet, so the remaining structural impediments to the Internet economy's growth are twofold: first, the digital isolation induced by the distance of Internet users from key telecommunications infrastructures; and second, the country's exposure to SMC outages.

... / ...

Keywords: NICT, submarine cables, infrastructures, telecommunications, firm performance

JEL codes: F02, O11, O33, O18, L25, L96

Acknowledgements

We would like to thank Raphaël Paris (Laboratoire Magma Volcan – UCA, CNRS) for the insightful discussions we had on seismic risk location. We also thank Simone Bertoli, Michaël Goujon, Mathilde Maurel, Jaime de Melo, and participants at the Ferdi-Banque de France research seminar, and at the DIAL 2017 and AFSE 2017 conferences, for their valuable comments. We are also grateful to Martine Bouchut (CERDI-UCA) for helping us in the programming of some variable calculation..

.../... We therefore adopt an instrumental variable (IV) approach reflecting these two sources of digital vulnerability. Exploiting the hierarchical structure of the World Bank Enterprise Survey dataset, multilevel IV estimations are conducted on a large sample of firms from more than 2,600 locations in some 60 developing and transition countries. They stress the large local impacts of an increase in the local incidence of email use by firms, induced by a lesser digital vulnerability, on a firm's average annual sales and sales per worker, and, to a lesser extent, on temporary employment. Estimated relationships are robust across a range of alternative sampling and specifications.

1- Introduction

Over the last few decades, international connectivity underwent a dramatic improvement promoted by the laying of 321 fibre submarine cables (SMC) over 1990–2015. Today, more than 99% of the world's telecommunications – Internet content, phone and video calls, classified diplomatic messages – passes through SMCs. And according to the OECD (Weller & Woodcock, 2013), in 2013, “20 households with average broadband usage generate as much traffic as the entire Internet carried in 1995”. SMC international networks now irrigate a USD 20.4 trillion industry and connect 3 billion Internet users across the world (Towela & Tesfaye, 2015).

Almost all coastal developing and transition countries are now plugged into the global Internet through SMCs. Fast-growing Asian and South American countries have been rapidly connected to Northern economies, and Africa's digital isolation from the rest of the world has rapidly fallen since 2005. This rapid expansion of the international broadband infrastructure network and the following boom in Internet services raises strong expectations for many low-income countries' economic catch-up, notably, through its potential for fostering innovation, productivity, trade, and job creation.

In the 2000s, a first strand of the literature started to study the macro-level effect of Internet penetration rates on international trade, in the wake of the “Internet bubble” and the financial collapse that followed. These studies generally find a positive effect of Internet access on trade in developing countries (Freund & Weinhold, 2002, 2004), especially on exports (Clarke & Wallsten, 2006) and, more specifically, service exports (Choi, 2010). Other macro-level studies have pointed out the contribution of Internet development to human capital and labour productivity (Cette et al., 2015). Notably, Choi and Yi (2009) assume that the Internet can facilitate knowledge spill-overs and may thus contribute to economic growth. Examining the impact of higher broadband penetration rates in 25 OECD countries, Czernich et al. (2011) get comparable results.

Recent studies have focused on the micro-level effect of Internet access and usage on firms' outcomes. For example, Clarke (2008) finds that the use of the Internet by firms in Eastern Europe and Central Asia has a significant impact on their decision to export but no effect on their intensive margin. Clarke et al. (2015) show that improved Internet increases firms' sales growth and labour productivity and that this effect is greater for small firms. In this line of research, Paunov and Rollo

(2015) find that, in 117 developing countries, the intensity of Internet use at the industry level has a significant and positive effect on labour productivity, especially in the most productive firms of the sample. In contrast to these studies, Chowdhury and Wolf (2003) find that ICT investment in three East African countries (Kenya, Tanzania, Uganda) has no direct significant effect on SME performance but a positive effect on general market expansion (measured by an index increasing with the output share sold far away from the local market). Although non-exhaustive, this short literature review points to a positive macro-level effect of Internet expansion on trade and growth in developing countries, supported at the micro-level by evidence on the growth in firms' sales, exports, and productivity.

This paper tries to bring additional insights into this area of research by providing new evidence on the local impact of improved access to the Internet on three firm outcomes – their revenue, their labour productivity, and their number of temporary employees – in developing and transition countries. To estimate this impact, we propose a novel identification strategy emphasizing the weaknesses of the maritime and terrestrial broadband infrastructure network. In fact, the worldwide deployment of SMCs is boosting the economy in many developing countries; but in parallel, it has reinforced the digital isolation of Internet users in countries facing difficulties expanding and managing large terrestrial infrastructure networks and has increased countries' exposure to SMC outages. These two dimensions of *digital vulnerability* – which could be defined as the risk for a country and its population to have their access to the global Internet hampered by failures in their telecommunications network – partly rely on exogenous factors, underlying the instrumental variable (IV) approach adopted in this paper.

How telecommunications “superhighways” impact the economy is not a new question (Röller & Wavermann, 2001). However, to our knowledge, few studies have examined the contribution of the recent deployment of broadband infrastructures to the growth process (Czernich et al., 2010), especially to development outcomes (Hjort & Poulsen, 2017; Tian, 2018). The empirical approach followed in these three papers highlights how terrestrial infrastructure network coverage, preceding the introduction of broadband technologies, is a good predictor of the current access to broadband Internet. Our identification strategy differs from this approach in that (i) comprehensive information on the terrestrial infrastructure network is missing for a large sample of developing and transition countries and (ii) the cross-sectional data from the World Bank Enterprise Surveys (WBES) used in this paper do not allow exploitation of the time dimension to identify exogenous internal variation in Internet access.

Instead, we construct a set of instrumental variables reflecting geographical and geological determinants of the SMC network's vulnerability and explaining cross-country and cross-location variations in Internet penetration rates. On the one hand, we approximate digital isolation by the geographic distance of Internet users from infrastructure nodes (SMC landing stations or Internet exchange points), building on the general rule stating that the cost of Internet communications is equal to the speed of the Internet times the distance covered (Weller & Woodcock, 2013). In other words, the longer the Internet user's distance to SMC landing stations, the stronger the need for

terrestrial infrastructure's deployment and maintenance, the higher the cost passed on Internet communications, and the lower the Internet speed (Nyirenda & Tesfaye, 2015, Bates, 2014; Weller & Woodcock, 2013). On the other hand, in various regions, the maritime infrastructure is exposed to the risk of multiple cable outages caused by seismic events in their neighbourhoods (Carter et al., 2009; Widmer et al., 2010; Palmer-Felgate et al., 2013). The vulnerability of the broadband infrastructure to SMC outages is indeed becoming a major issue for many developed and developing countries, inducing direct and indirect economic costs that amount to tens or hundreds of millions of dollars. Therefore, our identification strategy sheds light on the contribution of these sources of digital vulnerability to Internet penetration rates among firms, putting in evidence of an understudied dimension of the recent high-capacity broadband infrastructure expansion.

Estimations are conducted using WBES data on firms' characteristics, outcomes, and Internet usage aggregated at the location level. Data aggregation is motivated by two concerns. The first and most important concern is related to network externalities (Röller & Waverman, 2001) by which decisions of firms that are geographically close and carrying out similar activities are probably correlated with each other. Aggregating firm-level data at the location level therefore neutralizes possible endogeneity bias due to geographical network effects. The second concern is related to inference problems caused by the instrumentation of a dichotomous firm-level variable – the firm having used the Internet or not – in the first-stage equation, which lies between 0 and 1 once this variable is aggregated at the location level.

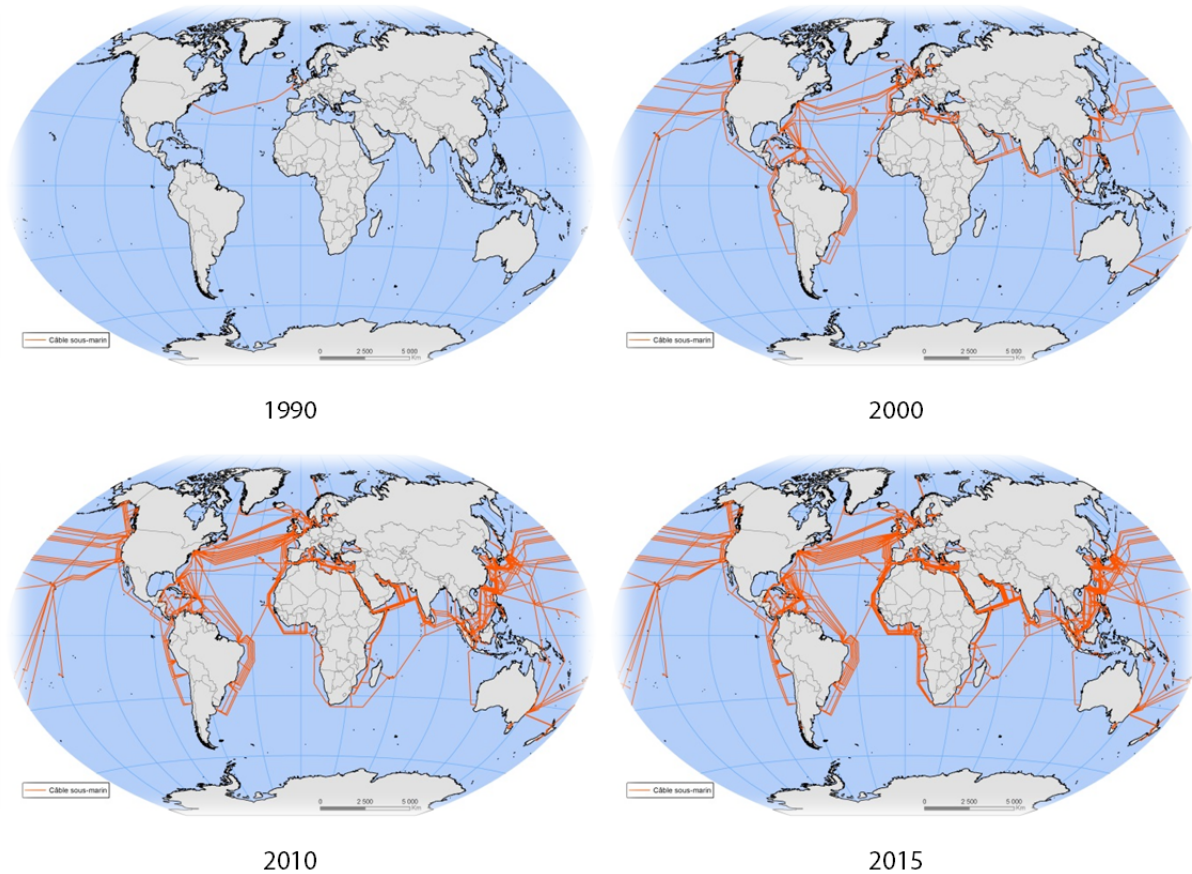
Based on a large sample of firms located in more than 2,600 cities/provinces in some 60 developing and transition countries, single-level and multilevel IV estimations stress that a lower digital vulnerability – induced by a lower digital isolation and a lower exposure to SMC outages – improves Internet penetration rates among firms and thereby boosts their revenue and labour productivity. Empirical evidence also points to a positive, softer, but less robust impact of Internet penetration rates on the number of full-time temporary employees.

Evidence is robust to the exclusion of exporters, big firms, foreign firms, and firms created after SMC arrival, and to the use of another proxy for firms' access to Internet. Consistent with the view that the telecommunications infrastructure has important externalities (Röller & Waverman, 2001), this paper emphasizes the large effects of broadband Internet at the location level. This evidence also supports that the digital divide, i.e., disparities in the quality of Internet access across different cities or provinces, dampens the effect of improved Internet access on economic outcomes at the macro level. It therefore suggests that the impact of broadband arrival is heterogeneous within countries.

The next section explains how telecommunications infrastructure deployment and digital vulnerability are related and details the instrumental variable setup. The third section sets the empirical framework and the fourth section presents the main results. The fifth section addresses

potential endogeneity bias and checks whether estimated relationships are robust to the use of alternative sampling and specifications. The last section concludes.

Figure 1. International digital connectedness through fibre SMCs



Source: Authors. **Data:** Telegeography.

2- Telecommunications infrastructure deployment and digital vulnerabilities

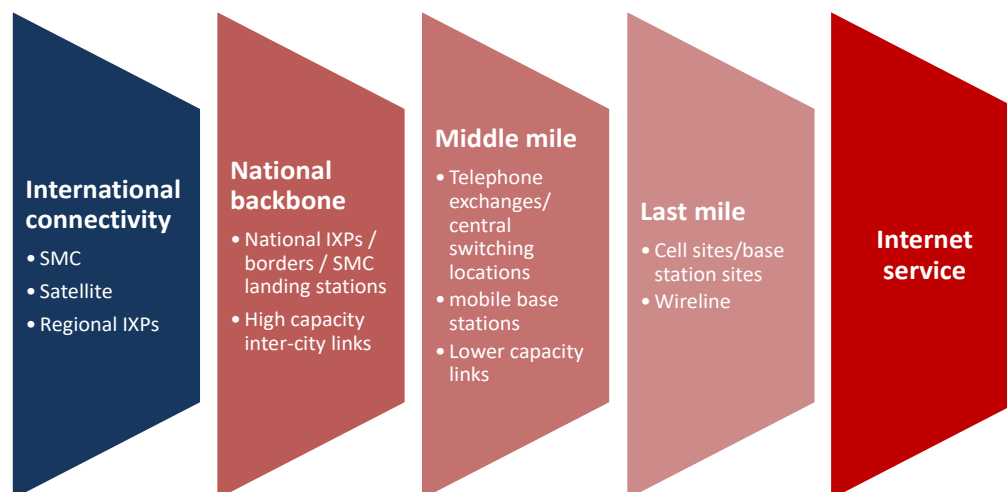
In this section, we set up a conceptual framework explaining how the deployment of broadband submarine and terrestrial infrastructures has boosted the Internet economy but has in parallel exposed developing economies to new sources of vulnerabilities: on the one hand, the digital isolation induced by the terrestrial distance between Internet users and existing broadband infrastructures; on the other hand, the network exposure to submarine cable outages. Building on these elements, we define an instrumental variable approach based on geographical and geological determinants of digital vulnerability.

2.1. Telecommunications infrastructure deployment and the Internet access value chain

International connectivity is the first element of the Internet access value chain, presented in Figure 2. In the absence of SMC, a country has two solutions to get international Internet connection: (i)

buying expensive and limited Internet bandwidth from a SMC-connected neighbouring country or (ii) buying expensive, often slow and limited Internet bandwidth from telecommunications satellites.¹

Figure 2. Internet access value chain and key telecommunications infrastructures



Source: Authors, adapted from Schumann and Kende (2013).

The laying of SMC infrastructures is therefore the first step towards access to the global Internet and a catalyst for subsequent investments in backbone, middle-mile, and last-mile terrestrial telecommunications infrastructures by making them more profitable (Schumann & Kende, 2013). The number of SMCs plugging countries into the global Internet is expected to boost the Internet economy by affecting the total bandwidth available to Internet users, the cost of Internet services, the competition environment, and a country's Internet redundancy when cable outages occur (Weller & Woodcock, 2013; Schumann & Kende, 2013; Telegeography, 2016).

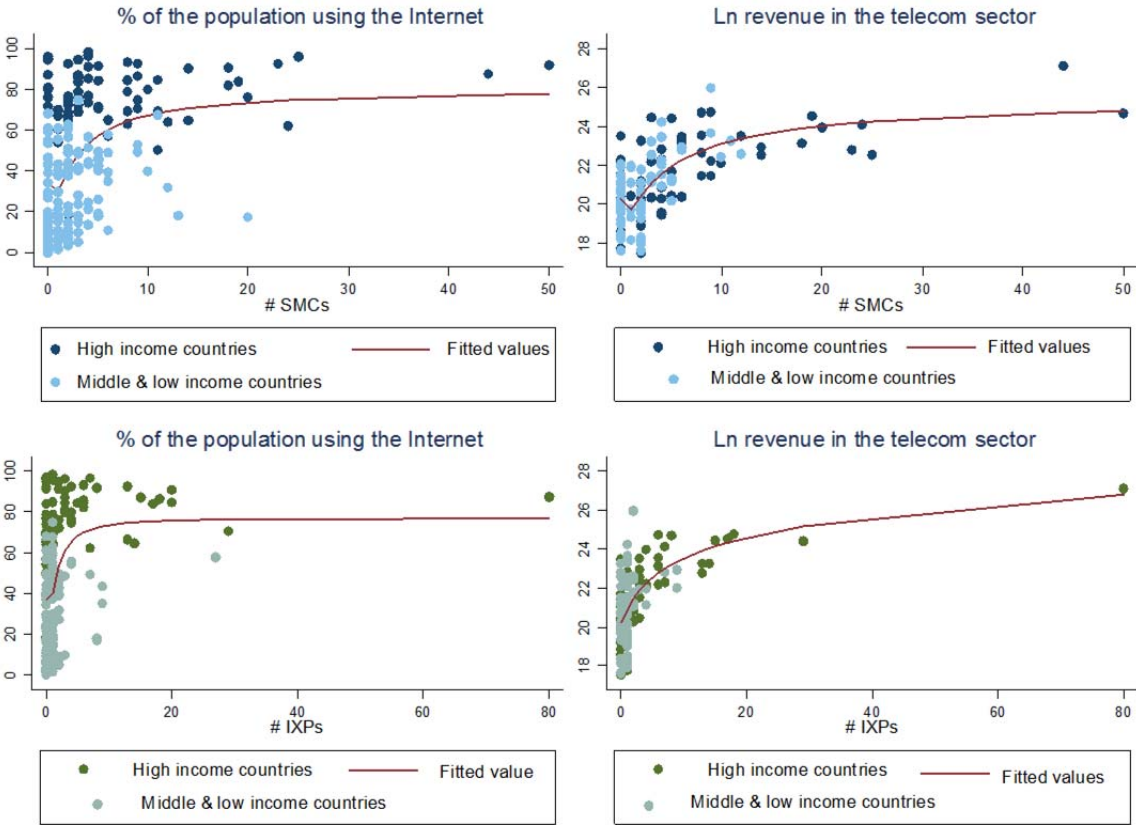
Once SMCs are plugged into the coast, Internet exchange points (IXPs) represent a key element of national and regional backbone telecommunications infrastructures and an important source of Internet bandwidth (Weller & Woodcock, 2013; OECD, 2014; Towela & Tesfaye, 2015). IXPs are national or regional Internet hubs where Internet protocol networks can exchange traffic with each other. IXPs allow latency to be reduced by favouring direct interconnections between countries and saving bandwidth for international communications by keeping local traffic locally. By doing this, IXPs also make it possible to carry Internet traffic at low cost, which reduces the cost of providing access to Internet. Therefore, IXPs increase Internet performance and cost-efficiency and are hence a central element of local and regional Internet ecosystems (Weller & Woodcock, 2013; OECD, 2014).

We highlight the contribution of SMCs and IXPs to the Internet economy in Graph 1. This graph shows a positive concave relationship between broadband infrastructures (alternatively measured

¹ Although the provision of new broadband Internet services by satellite is growing but still expensive.

by the number of SMCs and the number of IXPs) and Internet penetration rates on the one hand (left-hand side graphs) and the telecommunications sector’s revenue on the other hand (right-hand side graphs). This graphical evidence suggests that SMC and IXP deployment is good news for the telecoms industry and the whole Internet economy. However, the dispersion of observations around the fitted correlation lines also suggests that this relationship relies on other characteristics of the telecommunications infrastructure.

Graph 1. Correlation between SMC/IXP deployment and the Internet economy, worldwide evidence in 2014



Notes: Data on infrastructure deployment are drawn from Telegeography, Packet Clearing House, and Peering DB databases. Data on Internet penetration and revenue are drawn from the International Telecommunication Union (2016). **Sample:** Share of population using Internet variable: 201 developed and developing countries. All telecommunications revenue variable: 122 developed and developing countries.

2.2. Digital vulnerabilities

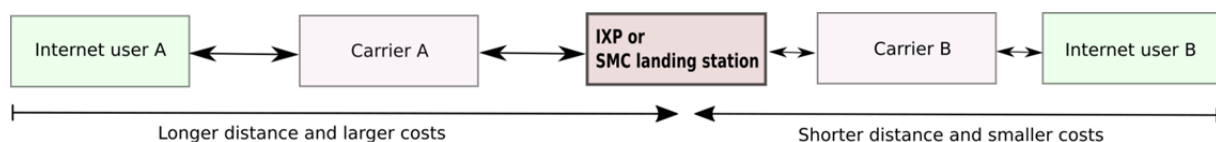
We define digital vulnerability as the risk of a country and its population to see their access to the global Internet hampered by failures in their telecommunications network. These failures may result from the telecommunications infrastructure network's incapacity to meet the population's needs (in terms of geographical coverage and Internet bandwidth), the infrastructure's gradual obsolescence, and its exposure to recurrent external shocks. These external shocks encompass telecoms infrastructure failures (network overload, server shutdown, SMC outages, IXP closure), power outages, political censorship, and computer attacks. Building on this definition, this paper emphasises the digital vulnerability resulting from telecommunications infrastructure failures: the digital isolation caused by missing or lacking terrestrial infrastructures and the country's exposure to SMC outages.²

2.2.1. Digital isolation and the infrastructure gap

In 2016, most developing and transition countries were plugged into the global Internet through SMCs, so the reduction of digital isolation through terrestrial infrastructure deployment is one of the greatest challenges for developing countries' Internet economies (Röller & Waverman, 2001). Low cross-border connectivity and missing national/regional terrestrial infrastructures are indeed the most important barriers to Internet access in low-income countries (Towela & Tesfaye, 2015; Bates, 2014). To stress the concern of digital isolation in the African context, Bates (2014, p. 2) stated that "fibre connectivity in cities that are far removed from submarine cable landing stations often costs five or six times as much as it does at the landing station".

Moreover, even in countries where the terrestrial infrastructure is not missing, a general rule according to which the cost of Internet communication equals the speed times the distance covered (Weller & Woodcock, 2013) makes the Internet user's average distance to the backbone infrastructure a critical determinant of digital isolation (figure 3). As a result, the longer the distance to backbone infrastructure nodes, the stronger the need for additional terrestrial infrastructure deployment and maintenance investments, the higher the cost passed on to Internet communications, the lower the Internet speed, and the greater the digital isolation.

Figure 3. Distance to key infrastructure nodes and Internet cost



Note: Adapted from Weller and Woodcock (2013). **Source:** Packet Clearing House.

² Putting aside the network's exposure to power outages, censorship, and computer attacks.

2.2.2. The country's exposure to SMC outages

A second digital vulnerability is the SMC network's exposure to cable outages. Between 2008 and 2012, data collected by Palmer-Felgate et al. (2013) show that 471 cable repairs were undertaken worldwide, geographically spread as follow: 186 repairs in the ACMA area (Atlantic Ocean), 115 in the MECMA area (Mediterranean Sea), 93 in the YZ area (Pacific Ocean, Asia side), 68 in the SEAIOCMA (2010-2011, Southeast Asia and Indian Ocean) and nine in the NAZ area (Pacific Ocean, North America side).³ As an illustration of economies' strong vulnerability to telecommunications SMC breaks, the simultaneous cut undergone by the SEA-ME-WE 3, SEA-ME-WE 4, and FLAG cables on December 19, 2008, induced a 24-hour outage and a 10-day slowdown of Internet access, affecting millions of people in Europe, the Middle East, South Asia, and even East Africa and causing a dramatic loss of Internet capacity for 14 countries.⁴

In addition to firms' direct costs of repairing damaged cables for telecoms operators, amounting to millions of dollars depending on cable repair frequency and length, there are indirect economic costs, rising to tens or hundreds of millions of dollars related to (Widmer et al., 2010; Clark, 2016)

- The reporting of repair costs on Internet tariffs and its consequences for Internet penetration;
- The rerouting of Internet traffic towards more expensive cable paths and its consequences on Internet capacity and tariffs;
- The disorganization of global manufacturing chains and Internet-related service provision (e.g., financial services).

Last but not least, these direct and indirect costs are amplified by delays in cable repairs. According to Palmer-Felgate et al. (2013), these delays vary significantly among maintenance areas and countries and are mainly caused by multiple outages caused by natural events such as earthquakes or typhoons, ships engaged in prior repairs (likely induced by multiple outages), repair permit acquisition delay, or operational issues (Borland, 2008).

2.3. Instrumental variable setup and descriptive evidence

This paper examines the causality of firms' Internet access on their performance, using as instrumental variables geographical and geological determinants of digital isolation and exposure to SMC outages.

³ Repair agreement areas are displayed in Appendix D.4.

⁴ According to France Telecom's press release dating from the 19th of December 2008 ("Three undersea cables cut: traffic greatly disturbed between Europe and Asia/Near East zone"), the following countries have suffered from dramatic Internet disruptions: Saudi Arabia: 55% out of service; Djibouti: 71% out of service; Egypt: 52% out of service; United Arab Emirates: 68% out of service; India: 82% out of service; Lebanon: 16% out of service; Malaysia: 42% out of service; Maldives: 100% out of service; Pakistan: 51% out of service; Qatar: 73% out of service; Syria: 36% out of service; Taiwan: 39% out of service; Yemen: 38% out of service; and Zambia: 62% out of service.

2.3.1. The need for telecommunications infrastructure deployment

As mentioned earlier, the Internet user's distance to the telecoms infrastructure network is a critical determinant of digital isolation. This observation is not new, as Röller and Waverman (2001) have already emphasized that in larger countries, telecommunications infrastructure investment requirements are higher. Telecommunications infrastructure deployment in large, landlocked, or fragmented territories is therefore more challenging, which constrains states in the provision of cost-efficient Internet services (Weller & Woodcock, 2013). Unfortunately, information on terrestrial backbone infrastructure deployment is not available for all developing and transition countries.⁵ However, information on the location of SMC landing stations and IXPs is available for most developing and transition countries. Building on this information, we proxy digital isolation at the country and location levels by two variables (Appendix A provides details on variables calculation), respectively:

- *the country's structural need for infrastructure deployment*, equal to the distance between the country's centroid and the closest SMC landing stations. This variable is intended to reflect geographical (and therefore exogenous) factors that will affect investments in infrastructure deployment and management required to bring the Internet in affordable conditions over the territory.
- *the firm's need for infrastructure deployment*, equal to the firm's distance from the closest infrastructure node (IXP or SMC landing station). We consider the firm's distance from infrastructure together with the country's structural need for infrastructures.⁶

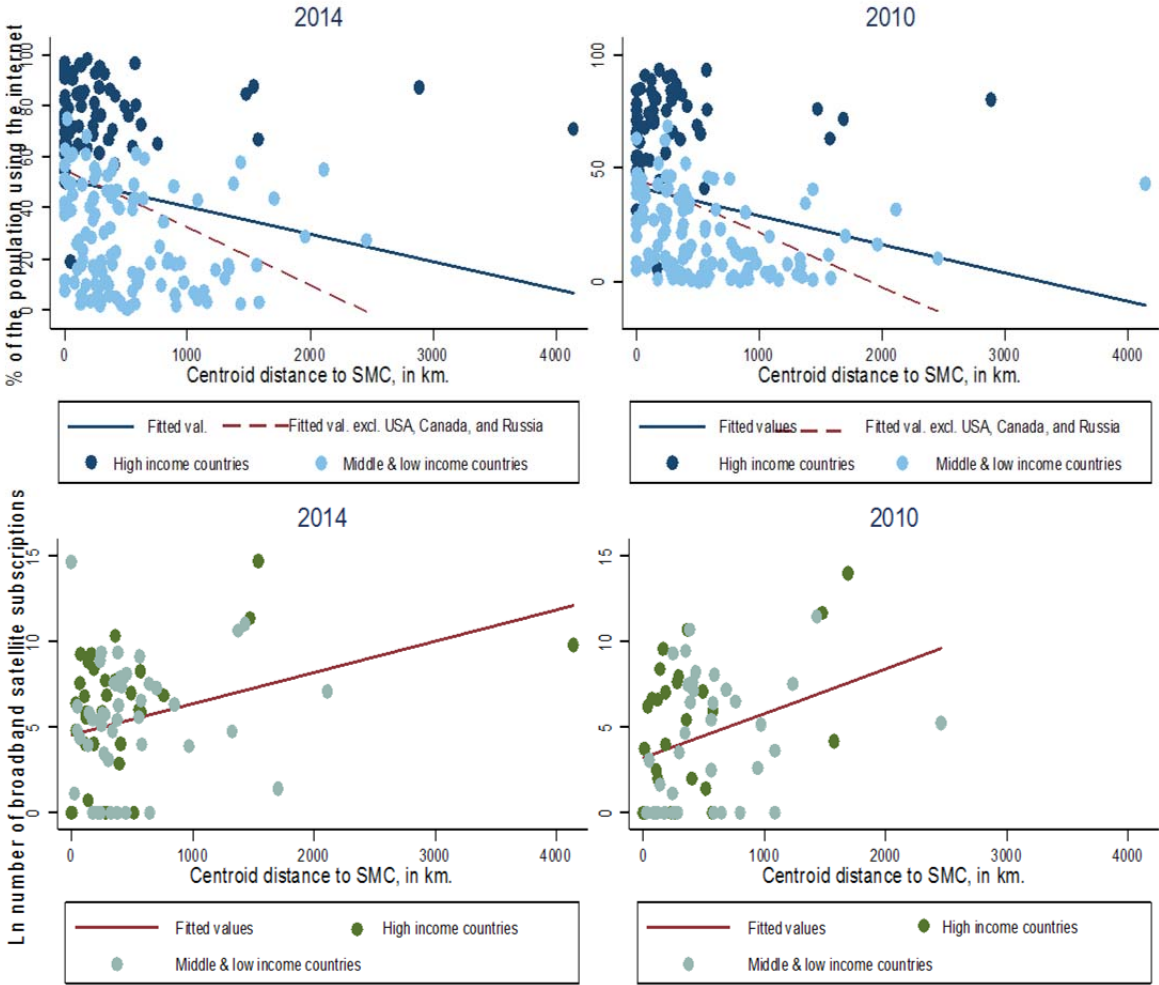
To illustrate the difference between the country's structural need and firms' need for infrastructures, Appendix D maps China's centroid, Chinese firms' location clusters, and China's IXPs and SMC landing stations' locations. Graph 2 below plots the 2014 and 2010 correlations between the structural need for telecoms infrastructures variable and the fibre Internet access variable: the share of the population using the Internet on the one hand, and the number of broadband satellite subscriptions on the other hand. Top graphs show a negative correlation between the centroids' distances to SMC landing stations and Internet penetration rates, while bottom graphs highlight a positive correlation between the centroids' distances to SMC and subscriptions to satellite broadband Internet, a more expensive alternative to fibre Internet access. This negative relationship between the centroid's distance to SMC and fibre Internet access finds additional evidence using averaged data on firms' Internet usage from the World Bank Enterprise Surveys (graph 3). Therefore, a greater structural need for infrastructure is at first sight negatively correlated with Internet access.

⁵ Open-source data on the backbone infrastructure in Africa are available at/and <http://www.africabandwidthmaps.com> (Hjort & Poulsen, 2017). We did not find this type of information for other developing and transition areas.

⁶ based on the observation that "the average distance between an exchange [point] and all of the members of the population it serves is a major measure of the efficiency, and thus the long-term success, of any exchange [point]" (Weller & Woodcock, 2013, p. 40).

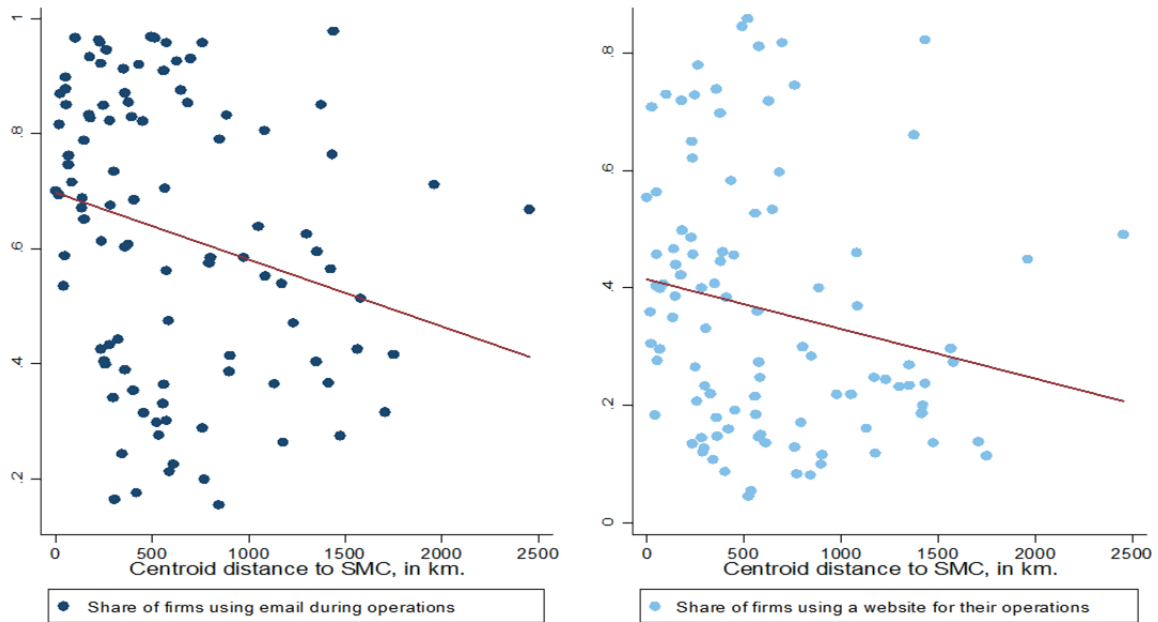
Graph 4 plots the WBES sample countries' averages of firms' distances to infrastructure nodes against the percentage of the population using the Internet (top graph), the country's share of firms using email during their operations (bottom left-hand side graph), or the country's share of firms using websites for their operations (bottom right-hand side graph). All graphs display a negative correlation between the average firm's infrastructure need and its inclination to use the Internet for its operations.

Graph 2. Correlation between countries' centroids' distances to SMC landing stations and Internet access, worldwide evidence in 2010 and 2014



Notes: High-income (dark-coloured) vs. middle- and low-income (light-coloured) countries. Top graphs: 192 countries (2014) and 188 countries (2010). Bottom graphs: 82 countries (2014), 66 countries (2010). The dashed fitted correlation line has been drawn to exclude outliers USA, Canada, and Russia. Data on Internet access proxies drawn from the International Telecommunication Union.

Graph 3. Correlation between countries' centroids' distances to SMC landing stations and firms' Internet usage, evidence from the WBES



Notes: averaged data from 94,750 firms located in 101 developing and transition countries, represented in the left-hand side graph, and from 94,616 firms located in 101 developing and transition countries, in the right-hand side graph. Firms were surveyed between 2006 and 2014.

Graph 4. Country correlation between firms' average distance to SMC/IXP and Internet use, evidence from ITU data and the WBES



Notes: Top graph uses data on the share of the population using the Internet in 106 developing and transition countries drawn from the International Telecommunication Union (ITU). Bottom graphs use averaged data from 96,840 firms (left) and 96,694 firms (right) from 106 developing and transition countries.

2.3.2. Infrastructures' exposure to seaquake-induced cable outages

Cable outages result from multiple factors that can be grouped into five sources of external shocks (Carter et al., 2009; Clark, 2016):

- Maritime activities (fishing nets, anchors), the main source of all cable faults;
- Natural events, such as earthquakes and seaquakes, volcanic eruptions, rising sea levels, and turbidity currents, causing in 2006 around 30% of cable breaks occurring in deep-sea water (probably more today, given the dramatic densification of the undersea cable network), and the main source of multiple cable breaks;
- Shark bites and whale entanglements, although the new generation of cables is much less prone to these risks;
- Piracy and sabotage.

Among exogenous causes of cable outages, earthquakes and seaquakes are the best documented ones. Seismic activity indeed may cause cable breaks directly or indirectly by provoking turbidities, landslides, and tsunamis (Soh et al., 2004; Carter et al., 2009; Clark, 2016). It is also likely to lengthen the time needed to repair cables by inducing simultaneous multiple outages (Palmer-Felgate et al., 2013). Last, because of their impact on the telecommunications sector's revenues, the risk of cable breaks induced by the recurrence of earthquakes and seaquakes is generally reported on telecommunications tariffs (Widmer et al., 2010).

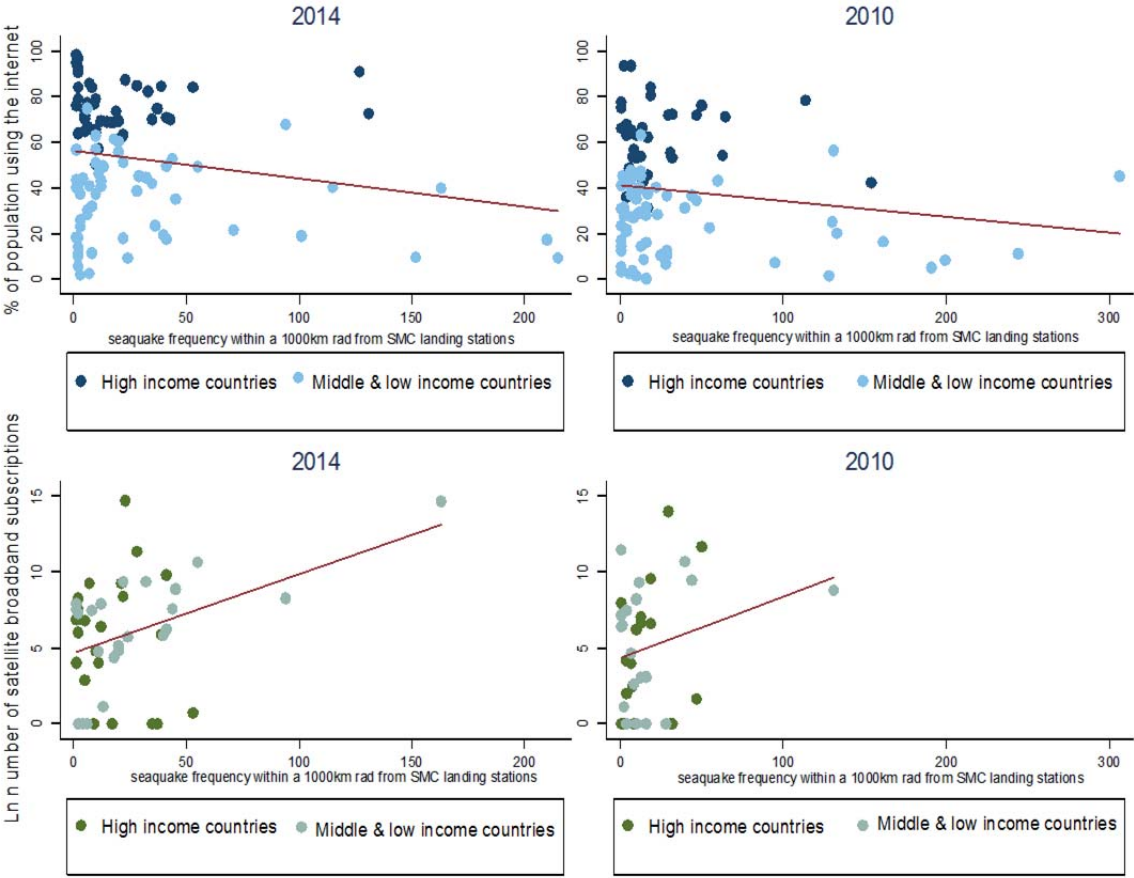
Because seismic activity may cause damage to the whole economy and therefore not only the SMC network, our identification strategy focuses more specifically on seaquake occurrence. We exploit information on the location, timing, frequency, and intensity of seaquakes to build a variable of a *country's exposure to seaquake-induced cable outages*: the annual frequency of seaquakes that occurred within a radius of 1000km from the SMC landing station. To respect identification restrictions, we only count seaquakes with intensity above 5 and exclude observations with seaquake intensity above 6.5 on the Richter scale: below 5, the seaquake might have little effect on the SMC infrastructure, while above 6.5, the seaquake may induce tsunamis that may damage entire coastal areas.⁷ Appendix D maps the digital vulnerability to seismic risk and the worldwide occurrence of seaquakes around SMC landing stations between 1990 and 2016.

Graph 5 plots the 2014 and 2010 correlations between the annual frequency of seaquakes within 1000km of SMC landing stations (of countries that experienced at least one seaquake above 5 on the Richter scale) and fibre Internet access proxies – i.e., the share of the population using the Internet and the number (in logarithm) of broadband satellite subscriptions. Top graphs show a

⁷ The lower bound has been chosen according to the work of Soh et al. (2004), who find that cable breaks occurred in the eastern part of Taiwan following earthquakes ranging from 5.0 to 6.0 on the Richter scale. The upper bound is based on an interview with Dr Raphaël Paris, Research Officer in volcanology at CNRS and Laboratoire Magmas et Volcans (LMV) (Observatoire de Physique du Globe de Clermont-Ferrand, Clermont-Auvergne University), who pointed out that the risk of tsunami becomes significant with seismic activity above 6.5 on the Richter scale.

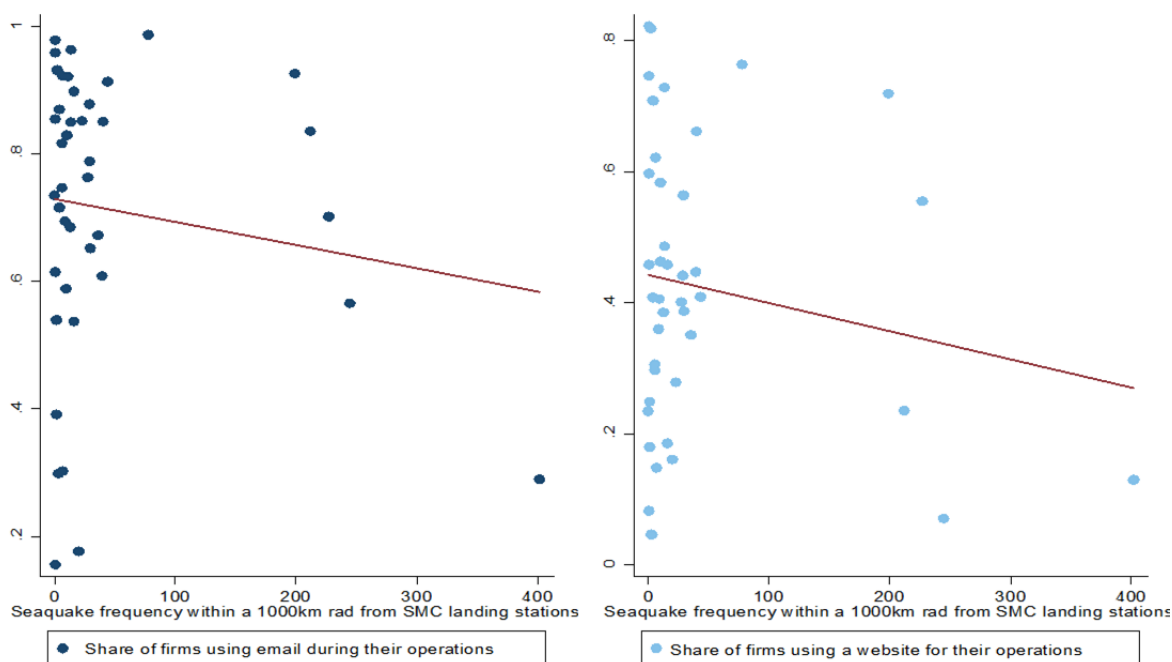
negative correlation between the countries' annual frequencies of seaquakes and Internet penetration rates, while bottom graphs highlight a positive correlation between the countries' annual frequencies of seaquakes and (more expensive) broadband Internet access through satellites. This evidence is corroborated by the graphical correlation between the countries' exposure to seaquake-induced cable faults and country averages of firm Internet use, displayed in graph 6. Therefore, a greater country exposure to seaquake-induced SMC outages is at first sight negatively correlated with Internet access.

Graph 5. Graphical correlation between annual seaquake frequency and Internet outcomes, worldwide evidence, 2010 and 2014



Notes: High-income (dark-coloured) vs. middle- and low-income (light-coloured) countries that experienced at least one seaquake above 5 on the Richter scale. Top graphs: 101 countries (2014) and 98 countries (2010). Bottom graphs: 46 countries (2014) and 38 countries (2010). Data on Internet access proxies drawn from the International Telecommunication Union.

Graph 6. Graphical correlation between annual seaquake frequency and firm Internet use, evidence from the WBES



Notes: Graphs exclude countries that did not experience seaquakes above 5 on the Richter scale within 1000km of their SMC landing stations. Averaged data from 55,595 firms located in 41 developing and transition countries represented in the left-hand side graph, and from 55,502 firms located in 41 developing and transition countries in the right-hand side graph. Firms were surveyed between 2006 and 2014.

3- Empirical framework

Our empirical approach aims to identify the local impact of Internet use on three firm outcomes: the firm's annual total sales, sales per worker, and number of temporary full-time employees. In this section, we describe the estimation strategy and the data.

3.1. Estimation strategy

3.1.1. Baseline estimation setup

Consider the following simple econometric model, which is the basis of pooled estimations conducted in this empirical analysis:

$$y_{i,t,j} = \gamma_0 + \gamma_1 Internet_{i,t,j} + \gamma_2 X_{i,t,j} + \gamma_3 V_{t,j} + \gamma_4 BBI_{t,j} + \pi_s + \mu_r + \tau_t + \varepsilon_{i,t,j} \quad (1)$$

where subscripts, i , t , and j respectively refer to the firm, the survey year,⁸ and the country, and where $y_{i,t,j}$, $Internet_{i,t,j}$, and $X_{i,t,j}$ are respectively variables of firm performance, firm Internet usage, and firm characteristics. Firm characteristics are the firm's size, age, and property status; the

⁸ Because the survey data used in this paper are not suited for panel analysis, pooled estimations are conducted so that two interview rounds in the same country are considered as distinct units of analysis.

size of the city where the firm operates; the firm's share of exports; the share of external funding; and the firm's electricity constraint. $V_{t,j}$ is a vector of key country-level controls: the population logarithm, the GDP per capita logarithm, a landlockedness dummy variable, democracy, and primary school enrolment. BBI_j is a set of country-level broadband infrastructure control variables: the number of SMCs, the number of IXPs, and the number of years since the first SMC laying. $\varepsilon_{i,t,j}$ is a random error term. We also include sector, region, and year-of-interview dummies (π_s, μ_r, δ_t). Firm-level variables are matched with country-level variables by country j and year of interview t .

3.1.2. Data aggregation and instrumental variable estimation framework

In a second step, the core of our analysis being the contribution of exogenous sources of digital vulnerability to Internet use and firm performance, we run pooled IV estimations based on the instrument set described in section 2. However, IV estimations using firm-level data face two important concerns. The first concern is related to network externalities (Röller & Waverman, 2001) by which firm-level decisions (including decisions to use the Internet or not) may be affected locally by other firms' decisions. In such context, orthogonality conditions may be violated, as our instruments for firms' email use vary at the country level and the location level. The second concern is related to the application of a two-stage IV regression framework with a firm-level dichotomous endogenous variable, Internet use, which may cause inference problems.

To address these issues, we conduct IV estimations averaging firm data at the location level, i.e., city or province level, so that (i) decisions of firms that are geographically close and carrying out similar activities are taken together, (ii) the dummy variable becomes a continuous variable lying between 0 and 1, and (iii) we account for network effects of Internet use among firms sharing geographical and sectorial characteristics.

We therefore average firm-level data following this ascending hierarchy⁹: firm's location (city or province), sector of activity, year of survey round, and country. This hierarchy in the data is justified by existing agglomeration externalities for firms operating in the same industry (Head et al., 1995, Litan & Rivlin, 2001). Therefore, clusters of firms from the same location but operating in different sectors, from different countries, and interviewed at different times represent distinct units of observation. Respecting this hierarchy in the data aggregation process increases the sample size and variability, helps avoid losing information on firms' characteristics and various statistical fallacies, and makes multi-level analysis possible (Hox, 2010; Stoker, 2010).

Aggregating the data at the location level, equation (1) becomes:

$$y_{l,s,t,j} = \delta_0 + \delta_1 Internet_{l,s,t,j} + \delta_2 X_{l,s,t,j} + \delta_3 V_{t,j} + \delta_4 BBI_{l,s,t,j} + \pi_s + \mu_r + \tau_t + \omega_{2l,s,t,j} \quad (2)$$

⁹ Based on the World Bank's sample stratification criteria. Aggregating information on the characteristics, Internet usage, and performance of heterogeneous firms geographically close but operating in different sectors and surveyed in different years would therefore be fallacious (Head et al., 1995; Litan & Rivlin, 2001; Hox, 2010; Stoker, 2010).

In addition to equation (1), the instrumental variable approach consists of estimating an additional first-stage equation:

$$Internet_{l,s,t,j} = \alpha_0 + \delta_1 Instruments_{l,s,t,j/t,j} + \alpha_2 X_{l,s,t,j} + \alpha_3 V_{t,j} + \alpha_4 BBI_{t,j} + \pi_s + \mu_r + \tau_t + \omega_{1l,s,t,j} \quad (3)$$

where additional subscripts *l* and *s* respectively refer to the firm's location *l* and sector *s* of activity. The $Instruments_{l,s,t,j/t,j}$ vector corresponds to instrumental variables described in section 2.3, and $\omega_{1l,s,t,j}$ and $\omega_{2l,s,t,j}$ are respectively the error terms of the first-stage and second-stage equations. Control variables are the same as those specified in equation (1). 2SLS and multilevel IV pooled estimations of equations (2) and (3) are conducted. In what follows, we detail the multilevel estimation framework.

3.1.3. Multilevel modelling

Multilevel research has shown that empirical analysis considering variables from different hierarchical levels as one single level leads to statistical and conceptual problems that could lead to the under-rejection of the null and to coefficient misinterpretation.¹⁰ Multilevel models take advantage of the hierarchical structure of the data to address these issues. Notably, the eventuality of intra-class correlation at various levels of the data (country, sector, or location levels) casts doubt on the unbiasedness of single-level estimations (Moulton, 1990; Hox, 2010), especially if we are interested in the local effect of aggregated variables. We therefore control for intra-class correlation by estimating equations (2) and (3) through a multi-level estimation framework (Hox, 2010).

For the sake of comparison, equation (1) is first estimated with disaggregated data within the following four-level random-effect model:

$$y_{i,l,s,t,j} = \Gamma_0 + \Gamma_1 Internet_{i,l,s,t,j} + \gamma_2 X_{i,l,s,t,j} + \gamma_3 V_{t,j} + \gamma_4 BBI_{t,j} + \pi_s + \mu_r + \tau_t + \varepsilon_{i,l,s,t,j} \quad (1')$$

where Γ_0 includes an intercept (γ'_0), a country-level random-intercept component ($\gamma''_{0,j}$), a sector-level random-intercept component ($\gamma'''_{0,s,j}$), and a location-level random-intercept component ($\gamma''''_{0,l,s,j}$), so that $\Gamma_0 = \gamma'_0 + \gamma''_{0,j} + \gamma'''_{0,s,j} + \gamma''''_{0,l,s,j}$. In addition to these random-intercept parameters, the model also includes country-level ($\gamma''_{1,j}$) and sector-level ($\gamma'''_{1,s,j}$) random components in the slope coefficient Γ_1 of the Internet variable, that is, $\Gamma_1 = \gamma'_1 + \gamma''_{1,j} + \gamma'''_{1,s,j}$.¹¹ The resulting "random-slope model" allows intercept and slope parameters to vary randomly across levels of the data and thereby to control for intra-class correlation between observation units at the country, sector, and location levels.

¹⁰ The statistical problem is that a standard statistical model applied to hierarchical data treats disaggregated values as independent information from higher-level subunits. The conceptual problem is related to (i) the ecological fallacy, i.e., when aggregated data are interpreted at the individual level; (ii) the atomistic fallacy, i.e., when conclusions at a higher level are based on inference made at a lower level; and (iii) the 'Simpson paradox', which stresses the risk of erroneous conclusions when data drawn from heterogeneous populations are collapsed and interpreted as if populations were homogeneous (Hox, 2010).

¹¹ Only significant random components are kept in the model. We did not add the location level in the random slope coefficient for computational reasons.

Second, using data averaged at the location level c , we conduct two-step IV estimations of equations (2') and (3'):

$$Internet_{l,s,j} = \Delta_0 + \Delta_1 Instruments_{l,s,j/j} + \delta_2 X_{l,s,j} + \delta_3 V_j + \delta_4 BBI_j + \pi_s + \mu_r + \tau_t + \varepsilon_{l,s,j} \quad (2')$$

$$y_{l,s,j} = A_0 + A_1 Internet_{l,s,j} + \alpha_2 X_{l,s,j} + \alpha_3 V_j + \alpha_4 BBI_j + \pi_s + \mu_r + \tau_t + \varepsilon_{l,s,j} \quad (3')$$

with the following random components: $\Delta_0 = \delta'_0 + \delta''_{0,j} + \delta'''_{0,s,j}$, $A_0 = \alpha'_0 + \alpha''_{0,j} + \alpha'''_{0,s,j}$, $\Delta_1 = \delta'_1 + \delta''_{1,j}$, and $A_1 = \alpha'_1 + \alpha''_{1,j}$. Since the standard errors of the second-stage equation (3') may be biased by the fact that $Internet_{c,j}$ is an estimated regressor, we implement a 100-replications non-parametric double-bootstrap of the first-stage and second-stage equations.

3.2. Data and measurement

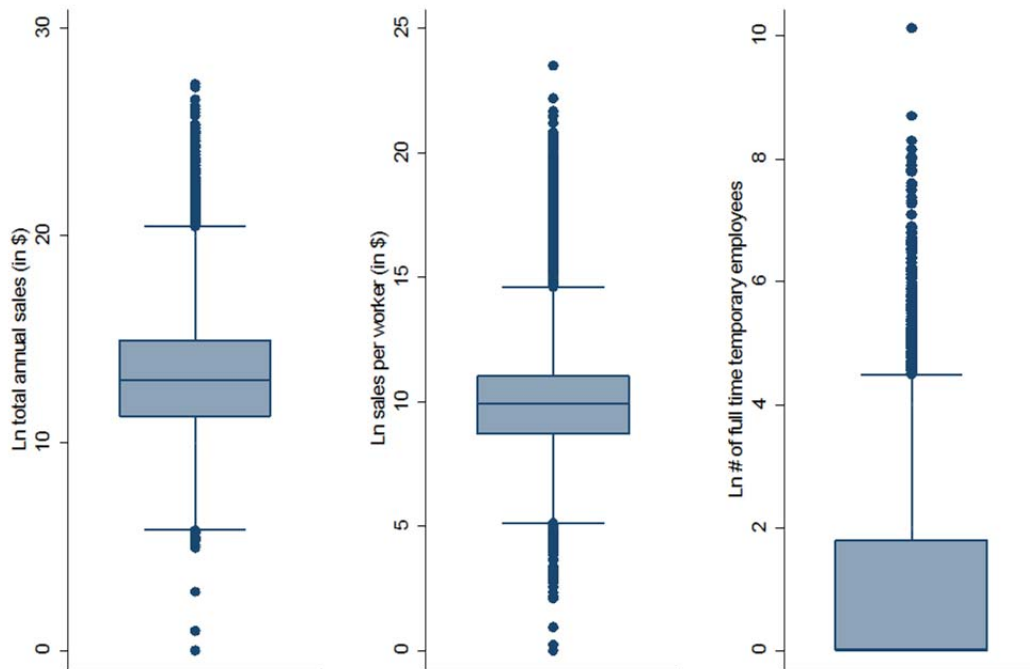
Appendix A explains the data and variables' definition, and Appendix B provides summary statistics of the variables listed below.

3.2.1. Dependent variables: Measuring firm's performance

All firm-level variables used in our model are drawn from the World Bank Enterprise Survey (WBES) cross-sectional dataset.¹² These data have been collected by the World Bank through random surveys conducted in a large sample of countries in firms in the manufacturing and service sectors. These surveys consist of interviewing business owners or top managers of formalized firms about several topics linked to their businesses and to the major obstacles they face to develop their activity. To measure individual firm performance, we alternatively use the logarithm of firm's total annual sales (converted into USD), total annual sales per full-time employee (converted into USD), and the number of temporary full-time employees. Graphs 7a and 7b plot the distribution of these variables. The average firm sales in the sample amount to 617,268 USD, the average sales per worker amount to 24,842 USD, and the average number of full-time temporary employees is equal to 2.6 workers.

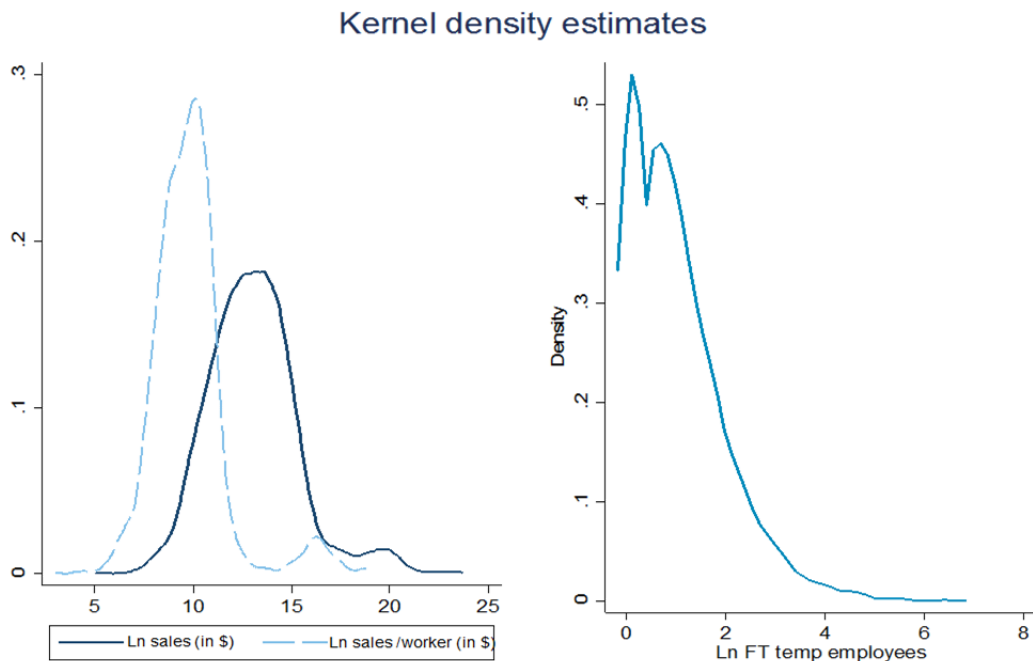
¹² Which survey design is not suited for panel data analysis (panel ID missing).

Graph 7a. Distributions of firm performance, firm-level data



Data: World Bank Enterprise Surveys. **Sales sample:** 26,488 observations, 62 countries. **Productivity sample:** 26,422 observations, 62 countries. **Employment sample:** 29,766 observations, 66 countries.

Graph 7b. Distributions of firm performance, location-level data



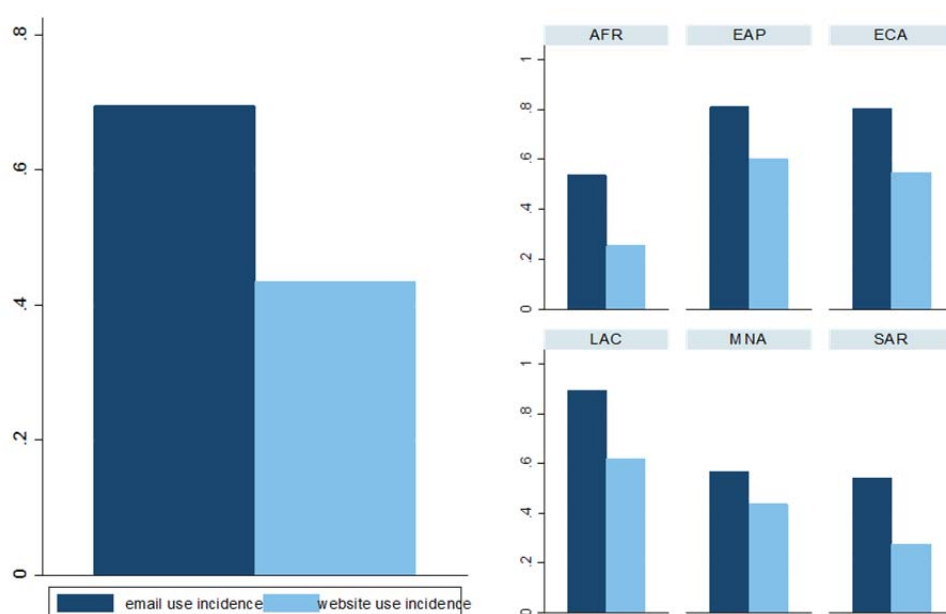
Data: World Bank Enterprise Surveys. **Sales sample:** 2,637 observations, 62 countries. **Productivity sample:** 2,634 observations, 62 countries. **Employment sample:** 2,800 observations, 66 countries.

3.2.2. Variables of interest: Internet use

Our main variable of interest (*Internet*) is a dummy variable equal to one if the firm *i* declares having used emails to communicate with its clients and suppliers during the past year, drawn from the World Bank Enterprise Surveys. This variable is our main variable of interest because email is the most basic way to use Internet, correlated with both simple and more complex uses of the Internet. To test the robustness of our results, we use as a dependent variable a dummy equal to one if the firm *i* declares having used a website to communicate with its clients and suppliers during the past year. This website-use variable is also of interest since it reflects a more specific and advanced use of the Internet by firms.

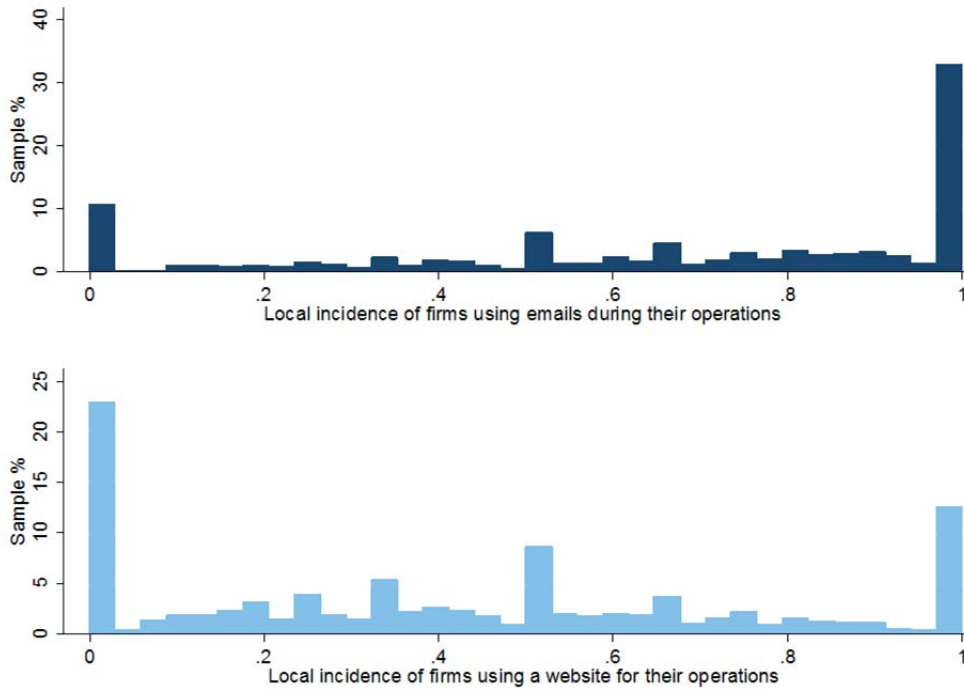
Graphs 8 and 9 below give some insights into firms' reported experience of Internet usage. While more than 70% (40%) of firms have reported using email (a website) during their operations, data aggregation at the location level permits us to get a continuous Internet penetration variable (lying between 0 and 1) and to get more variability in the sample.

Graph 8. Global incidence of Internet use by firms (global and region averages)



Data: World Bank Enterprise Surveys. Sample: 29,766 observations, 66 countries. AFR: Sub-Saharan Africa; EAP: East-Asia and Pacific; ECA: Eastern Europe and Central Asia; LAC: Latin America and Caribbean; MNA: Middle East and North Africa; SAR: South Asia Region.

Graph 9. Local incidence of Internet use by firms (location-level penetration rates)



Data: World Bank Enterprise Surveys. **Sample:** 3,372 observations in 66 countries.

3.2.3. Broadband infrastructure and instrumental variables

As explained in subsection 2.3, we instrument a firm's access to Internet using three instrumental variables: (i) the centroid's distance to the closest SMC landing station (country-level instrument), a variable reflecting the structural need for telecoms infrastructures; (ii) the firm's distance to the closest telecoms infrastructure node – either SMC landing station or IXP – a variable reflecting the firm's need for infrastructures (location-level instrument); and (iii) the annual frequency of seaquakes (lying between 5 and 6.5 on the Richter scale) occurring within 1000km of SMC landing stations (country-level instrument), reflecting the country's exposure to seaquake-induced cable outages.

To avoid omitted variable bias, we control for features of the telecommunications infrastructure network (*BBIn*) that are related to policy or to the maturity of the Internet economy. We therefore include as control variables *the number of SMCs* and *IXPs* and *the number of years since the first fibre SMC's arrival*. The importance of controlling for these three dimensions of the telecommunications infrastructure network to avoid omitted variable bias has been stressed in section 2. Moreover, beyond the contribution of SMCs and IXPs to Internet access, telecommunications infrastructure investment may lead to growth directly because their deployment stimulates economic activity, but also indirectly by creating new productive activities (Röller & Waverman, 2001).

4- Empirical results

We first proceed to a standard single-level empirical analysis of the contribution of the Internet to firm performance by conducting OLS and 2SLS-IV estimations of equations (1) and (2). Then, we adopt a multilevel framework to account for intra-class correlation and to control for unobserved specific random effects at various levels of the data.

4.1. Single-level empirical analysis

4.1.1. Pooled OLS estimates

Pooled single-level OLS estimates of equation (1) are reported in Table 1. They highlight that the use of email during firms' operations is positively and consistently correlated with firms' total sales and productivity but uncorrelated with temporary employment. BBI variables display low correlations with firm performance variables, especially when the data are averaged at the location level. Among these variables, the number of years since the first SMC's laying is the only variable consistently correlated with firm performance: positively with firms' total sales and productivity and negatively with temporary employment.

To ensure that the coefficient estimates of email use are not upward biased by unobserved country-level heterogeneity (Röller & Waverman, 2001; Wallsten, 2005), we report in Appendix C.2 (columns (1), (3), and (5)) estimates of benchmark OLS regressions replacing country-level variables with country dummies. The resulting estimates of email use coefficients are consistent and remain stable, suggesting that our model is correctly fitted.¹³ Their magnitude suggests that a 10% increase in Internet penetration among firms leads to a 7% increase in their revenue and productivity.

These OLS estimates provide first insights into the relationship between Internet infrastructures, Internet use, and firm performance. Because these estimated relationships may suffer from omitted variable bias or reverse causality from firm's outcomes to Internet use (Röller & Waverman, 2001; Czernich et al., 2011), we proceed to IV estimations in the next subsections.

¹³ However, coefficient estimates of the landlockedness and GDP per capita variables are respectively positive and negative, suggesting that the model could be misspecified. This coefficient turns negative in multilevel estimations conducted in the next sections (section 4.3, table 3), supporting that these unexpected correlations are driven by lower-level unobserved characteristics and justifying the use of a multilevel framework.

Table 1. Email use, broadband infrastructures, and firm performance: OLS single-level estimations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
				<i>Location level</i>				<i>Location level</i>				<i>Location level</i>
	Total sales				Sales/worker				Temporary employment			
Email use	1.252*** (0.112)	1.215*** (0.110)	0.530*** (0.128)	0.764*** (0.165)	0.558*** (0.0885)	0.540*** (0.0861)	0.426*** (0.113)	0.672*** (0.143)	0.259*** (0.0499)	0.0512 (0.0734)	0.0356 (0.0733)	-0.0320 (0.142)
BBI controls												
# SMC		0.436*** (0.0931)	0.141 (0.0933)	0.0653 (0.0574)		0.438*** (0.0930)	0.142 (0.100)	0.0614 (0.0605)		0.00988 (0.0206)	0.0468* (0.0282)	0.0624 (0.0465)
# IXP		0.0125 (0.00803)	0.0197* (0.0109)	-0.0107 (0.0111)		-0.0191*** (0.00734)	0.0182* (0.0103)	-0.0138 (0.00998)		0.0410*** (0.00525)	0.0385*** (0.00677)	0.0537*** (0.0114)
Time of 1st cable		-0.00711 (0.0168)	0.104*** (0.0204)	0.0812*** (0.0199)		-0.00770 (0.0166)	0.106*** (0.0208)	0.0879*** (0.0199)		-0.00317 (0.00516)	-0.0275*** (0.00774)	-0.0372*** (0.0135)
Macro controls												
Landlocked	0.218* (0.129)	0.767*** (0.133)	1.373*** (0.201)	1.036*** (0.183)	0.174 (0.129)	0.640*** (0.130)	1.262*** (0.203)	0.972*** (0.172)	-0.0145 (0.0463)	0.0670 (0.0477)	-0.0787 (0.0675)	-0.00440 (0.0981)
GDP per cap.	0.427*** (0.0691)	0.111 (0.0826)	-0.353** (0.144)	-0.210 (0.162)	0.420*** (0.0687)	0.0552 (0.0798)	-0.413*** (0.144)	-0.228 (0.162)	-0.160*** (0.0214)	-0.0945*** (0.0205)	-0.0615** (0.0267)	-0.0557 (0.0434)
Population	0.0298 (0.0453)	-0.157*** (0.0522)	-0.149*** (0.0476)	-0.144*** (0.0502)	-0.0503 (0.0414)	-0.210*** (0.0525)	-0.183*** (0.0513)	-0.146*** (0.0479)	0.0199 (0.0164)	-0.0180 (0.0165)	0.0451* (0.0245)	0.0344 (0.0349)
Education	0.000925 (0.00550)	-0.00669 (0.00596)	0.00416 (0.00485)	-0.00681 (0.00498)	-0.00450 (0.00525)	-0.0128** (0.00576)	-0.00183 (0.00474)	-0.0106** (0.00478)	-0.000194 (0.00180)	0.00127 (0.00155)	-0.00260 (0.00186)	0.00109 (0.00311)
Democracy	-0.0163 (0.0108)	-0.0232** (0.0102)	-0.0231 (0.0161)	-0.00127 (0.00933)	-0.00338 (0.0113)	-0.00708 (0.0109)	-0.0224 (0.0171)	0.00417 (0.00959)	0.00227 (0.00434)	-0.00206 (0.00400)	0.00606 (0.00558)	0.00470 (0.00833)

Micro controls												
State-owned	-0.242 (0.215)	-0.252 (0.286)			-0.292* (0.158)	-0.262 (0.247)			0.412** (0.166)	-0.0317 (0.215)		
Foreign	0.409*** (0.108)	-0.0527 (0.343)			0.396*** (0.145)	0.197 (0.231)			0.0235 (0.124)	0.0405 (0.275)		
Age	0.103* (0.0592)	0.0830 (0.122)			0.0259 (0.0598)	-0.00744 (0.0974)			-0.0302 (0.0271)	-0.0506 (0.0973)		
City size	-0.0536 (0.0420)	-0.142*** (0.0456)			-0.0404 (0.0416)	-0.119*** (0.0432)			-0.0112 (0.0195)	-0.0395 (0.0368)		
Large size	2.909*** (0.0980)	3.258*** (0.241)			-0.0122 (0.0958)	0.100 (0.202)			1.052*** (0.0917)	0.682*** (0.217)		
Medium size	1.320*** (0.0809)	1.616*** (0.209)			-0.0274 (0.0766)	0.116 (0.174)			0.433*** (0.0456)	-0.0874 (0.140)		
External funding	0.00937*** (0.00323)	0.00704* (0.00391)			0.00837*** (0.00306)	0.00986*** (0.00250)			0.00238** (0.00115)	0.00295 (0.00294)		
Electricity obstacle	0.00464 (0.0239)	-0.000809 (0.0760)			-0.0146 (0.0253)	0.00877 (0.0647)			0.0221 (0.0154)	0.0255 (0.0483)		
% of exports	0.00231* (0.00123)	-0.00200 (0.00287)			0.000589 (0.00119)	-0.00482 (0.00297)			-0.000273 (0.00101)	0.00313 (0.00395)		
Dummies	Region, year of interview, sector, firm size				Region, year of interview, sector, firm size				Region, year of interview, sector, firm size			
<i>N</i>	59,418	59,184	33,394	3,372	59,230	59,000	33,315	3,369	68,924	68,619	38,304	3,560
<i>Countries</i>	86	84	67	67	86	84	67	67	90	88	71	71
<i>R</i> ²	0.202	0.225	0.451	0.645	0.172	0.209	0.211	0.395	0.083	0.092	0.184	0.309

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses. Standard errors are robust to heteroscedasticity, are clustered by country-year-sector-city in OLS estimations using firm-level data, and are clustered by country-year-sector in OLS estimations using location-level data.

4.1.2. 2SLS-IV estimations

Results of 2SLS-IV estimations are presented in Table 2. First-stage estimates show that the stronger the firm's need for infrastructure, i.e., the further the firm from the cable landing point, the weaker the likelihood of the firm's using the Internet and the lower its performance. We also find a negative and significant effect of the structural need for infrastructure, i.e., the centroid's distance to an SMC landing point, on email use. Last, the exposure to seaquake-induced cable faults, proxied by the frequency of seaquakes within a 1000km radius, has, as expected, a negative and significant effect on email use.

Second-stage estimates support a stronger positive and 1%-significant effect of email use on annual sales, productivity, and temporary employment. According to our estimates, a 10% increase in email use among the population of firms raises more than proportionally the average logarithm of firms' annual sales by 24%, and by 22% the average logarithm of sales per full-time employee (corresponding approximately to a one-standard-deviation increase in these outcomes). Identification tests suggest that the instrument set is valid in the sales and productivity equations. Regarding the employment equation, the Hansen test rejects the hypothesis of residual orthogonality at a 10% confidence level, which casts doubt on the reliability of the estimated coefficient.

To ensure that a general "growth effect" of infrastructures driven by unobserved country-specific factors does not bias coefficients upward (Röller & Waverman, 2001), we conducted benchmark IV estimations using the firm distance to infrastructure as the only instrument and replacing country-level variables with country dummies. Estimates, reported in Appendix C.2 (columns (1), (2), and (3)), confirm the significant effect of email use on a firm's revenue and productivity, despite a single-instrument set's being somewhat weaker.

Table 2. Email use, digital vulnerabilities, and firm performance – IV estimations

	(1)	(3)	(6)
	Total sales	Sales/worker	Temporary employment
2nd stage estimates			
Email use	2.394** (0.964)	2.258** (0.917)	1.746** (0.878)
BBI, macro, and micro controls	Yes	Yes	Yes
1st stage estimates			
Instruments:			
Firm distance to SMC/IXP	-0.015** (0.005)	-0.015** (0.005)	-0.012** (0.005)
Centroid distance to SMC	-0.060*** (0.019)	-0.060*** (0.019)	-0.043** (0.017)
Seaquake freq., 1000km rad.	-0.003* (0.0018)	-0.003** (0.0018)	-0.003 (0.0022)
BBI, macro, and micro controls	Yes	Yes	Yes
Dummies	Region, year of interview, sector, firm size		
N	2633	2633	2797
Countries	62	62	66
Hansen test (p. value)	0.75	0.34	0.10
Weak ident. test (F-stat)	11.54***	11.50***	7.16***
Under-ident. test (LM-stat)	27.38***	27.32***	18.73***

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses. Standard errors are robust to heteroscedasticity and are clustered by country-year-sector. Firm-level variables are averaged at the country-year-sector-location level. Standard errors are robust to heteroscedasticity and are clustered by country-year-sector.

Until now, the hierarchical structure of the data and the eventuality of intra-class correlation in various layers of the data have been not accounted for in OLS and 2SLS estimations. In the next subsection, we estimate equations (1), (2), and (3) by implementing the multi-level empirical analysis explained in subsection 3.1.3.

4.2. Multilevel analysis

4.2.1. Multilevel estimations of equation (1)

Table 3 reports four-level and three-level estimations of equation (1') – performed on firm-level and location-level data, respectively – and three-level IV estimations of equations (1') and (2') – performed on location-level data. According to the four- and three-level estimates of equation (1'), a 10% increase in email use is significantly associated with an 8% increase in firms' sales, a 7% increase in labour productivity, and a 2–3% increase in the number of temporary employees.

First, the comparison between the coefficient estimates of equation (1) in table 1 and those in table 3 (columns (1), (2), (4), (5), (7), and (8)) stresses that the relationship between Internet use and firm

performance seems to be understated in single-level estimations. Second, the comparison of estimates in table 3 with coefficient estimates of the country-dummy three-level model reported in Appendix C.2 (columns (2), (4), and (5)) highlights the convergence of estimated coefficients across different multi-level models. This comparison permits us to confirm that estimated coefficients are not biased upward because of a “growth effect” of telecoms infrastructure (Röller & Waverman, 2001). Multi-level estimates of equation (1) in appendix C.2 (column (6)) also confirm the positive, consistent, and significant correlation between Internet use and the firm’s number of temporary employees, as reported in table 3.

4.2.2. Three-level IV estimations

Next, we run nonparametric double-bootstrapped three-level IV estimations of equations (2') and (3') and report second-stage estimates in table 3 (columns (3), (6), and (9)). First-stage three-level (single-bootstrapped) estimates are reported in Appendix C.1 and stress the significant negative contribution of the country’s exposure to seaquake-induced SMC outages to Internet access and the negative contribution of the firm’s distance to infrastructure nodes, but the nuanced contribution of the centroid’s distance to SMCs.¹⁴

Second-stage estimates show that a 10% increase in the incidence of email use among firms increases by 26% their revenues, by 21% their productivity, and by 6% temporary employment. These estimated impacts of email use on firms’ sales and productivity, reported in columns (3) and (6), are therefore very close to those provided by the single-level IV-2SLS estimator in table 2. Both 2SLS and three-level IV estimates therefore stress that the relationship between email use and firm performance resulting from equation (1)’s estimations tends to be strongly understated.

Interestingly, the estimated impact of email use on temporary employment in column (9) is less than half of that provided by the IV-2SLS estimator. Therefore, this evidence, combined with the previous evidence of a significant positive correlation between email use and employment found in multi-level estimations of equation (1), suggests that the rejection of the orthogonality conditions in IV-2SLS estimations of the employment equation in table 2 (column (3)) might be explained by the presence of intra-class correlation at the country and/or sector levels.

¹⁴ Hox (2010) recommends at least 1000 bootstrap samples for sufficient accuracy. However, due to computational constraints, we limit the double-bootstrap samples to 100 replications and find that non-bootstrapped IV estimations were actually understating the standard error’s size. This suggests that rejections of the null should not be affected by a smaller replication sample.

Table 3. Multi-level estimations

Dep. Var.:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total sales			Sales/worker			Temporary employment		
	4-level Eq(1)	3-level Eq(1)	Bootstrapped 3-level IV	4-level Eq(1)	3-level Eq(1)	Bootstrapped 3-level IV	4-level Eq(1)	3-level Eq(1)	Bootstrapped 3-level IV
Email use	0.849*** (0.0523)	0.841*** (0.115)	2.639*** (0.259)	0.661*** (0.049)	0.682*** (0.113)	2.061*** (0.229)	0.203*** (0.0234)	0.358*** (0.0744)	0.636*** (0.160)
<i>BBI controls</i>									
# SMC	-1.539*** (0.175)	-1.773*** (0.207)	-1.428*** (0.170)	-1.544*** (0.187)	-1.781*** (0.185)	-0.459*** (0.138)	-0.0008 (0.0301)	0.0251 (0.0338)	-0.0288 (0.0440)
# IXP	1.154*** (0.192)	0.684*** (0.198)	-0.458*** (0.111)	1.123*** (0.194)	0.722*** (0.183)	-0.230* (0.138)	0.0120 (0.0103)	-0.00380 (0.0113)	0.0621 (0.0511)
Time of 1st cable	-0.326*** (0.123)	-0.0541 (0.112)	-0.218** (0.103)	-0.316** (0.127)	-0.0888 (0.112)	-6.971*** (2.366)	-0.00300 (0.0122)	-0.0139 (0.0107)	0.0241 (0.0161)
<i>Macro controls</i>									
Landlocked	-4.753*** (1.688)	-2.610** (1.181)	-2.287*** (0.376)	-4.981*** (1.760)	-2.959** (1.221)	-2.551*** (0.471)	-0.0191 (0.105)	-0.164 (0.148)	-0.245** (0.101)
GDP per cap.	-2.706** (1.074)	0.204 (0.508)	0.891*** (0.230)	-3.017*** (1.166)	-0.186 (0.562)	0.638* (0.352)	-0.138*** (0.0534)	-0.165*** (0.0574)	-0.184*** (0.0659)
Population	0.800 (0.614)	0.347 (0.341)	0.934*** (0.157)	0.859 (0.633)	0.406 (0.372)	0.991*** (0.222)	-0.0216 (0.0355)	-0.0153 (0.0492)	-0.0725 (0.0539)
Education	0.0201 (0.0266)	-0.0172 (0.0273)	-0.0236** (0.0108)	0.0169 (0.0271)	-0.00791 (0.0281)	-0.0183 (0.0133)	-0.00150 (0.00217)	0.00248 (0.00255)	0.000615 (0.00238)
Democracy	0.0938 (0.101)	-0.0374 (0.0793)	0.117** (0.0573)	0.0853 (0.106)	0.0145 (0.0799)	0.176* (0.0900)	0.00157 (0.00995)	-0.000487 (0.0116)	-0.0150 (0.00987)
<i>Firm controls</i>									
State-owned	0.276*** (0.0893)	0.373* (0.199)	0.333 (0.218)	0.114 (0.102)	0.326 (0.222)	0.323 (0.212)	0.187 (0.138)	0.252 (0.257)	0.195 (0.218)
Foreign	0.721*** (0.0565)	0.601*** (0.180)	0.181 (0.178)	0.544*** (0.0485)	0.410*** (0.139)	0.105 (0.146)	0.0986* (0.0532)	0.0610 (0.136)	0.0170 (0.119)
Age	0.180*** (0.0163)	0.142** (0.0593)	0.117* (0.0702)	0.0760*** (0.0187)	0.0296 (0.0517)	0.0347 (0.0554)	0.0108 (0.0144)	-0.0164 (0.0395)	-0.0345 (0.0385)
City size	-0.0183 (0.0160)	-0.0398 (0.0299)	0.0416 (0.0300)	-0.00876 (0.0157)	-0.0334 (0.0232)	0.0339 (0.0263)	0.0154 (0.0138)	0.0217 (0.0224)	0.0375* (0.0223)
External funding	0.00582*** (0.000870)	0.00738*** (0.00182)	0.00629*** (0.00187)	0.00436** (0.0007)	0.00557*** (0.00195)	0.00425** (0.00204)	0.00276*** (0.000519)	0.00542*** (0.00122)	0.0064*** (0.0015)
Electricity obst.	-0.0110 (0.0106)	-0.0221 (0.0318)	-0.0447 (0.0324)	-0.0167* (0.00955)	-0.0225 (0.0294)	-0.0374 (0.0323)	0.0162*** (0.00585)	0.0379** (0.0182)	0.0161 (0.0235)
% exports	0.00246*** (0.000740)	0.00330* (0.00173)	0.00244 (0.00162)	0.0009 (0.0007)	0.00126 (0.00151)	0.000725 (0.00150)	0.00237*** (0.000612)	0.000965 (0.00150)	0.0009 (0.0015)
Observation unit	Firm	Location	Location	Firm	Location	Location	Firm	Location	Location
Dummies	Region, year of interview, sector, firm size								
Random intercepts	Country, sector, location	Country, sector	Country, sector	Country, sector, location	Country, sector	Country, sector	Country, sector, location	Country, sector	Country, sector
Random slope [email use]	Country	Country	Country, sector	Country	Country	Country, sector	Country	Country	Country, sector
N	33,442	3,380	2,637	33,363	3,377	2,630	38,355	3,567	2,592
Countries	67	67	62	67	67	62	71	71	66
Completed replications			100			100			100

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses and are robust to heteroscedasticity. Only significant random components have been kept in multi-level estimations. Three-level IV estimates result from a 100 replications double-bootstrap.

Last, to ensure that the three-level IV estimates are not biased upward because of country-specific effects, we again conduct three-level benchmark IV estimations using the firms' distance to infrastructure as the only instrument and replacing country-level variables with country dummies. The estimates, reported in Appendix C.3, show a twice as significant effect of email use on a firm's revenue and productivity but no significant effect on temporary employment.

In this section, we addressed potential endogeneity bias through an IV approach highlighting the role of structural determinants of digital vulnerability through a multi-level modelling of the data and by comparing estimates with those resulting from country-dummy variable estimations reported in appendix C. IV estimations shed light on large, positive, local effects of an increased access to the Internet on firm performance. Notably, sales and productivity equations' estimates appear consistent across different models, which is less the case of the relationship between email use and temporary employment. In the next section, we proceed to additional robustness tests.

5- Robustness analysis

5.1. Sample restrictions and the reverse causality concern

As first robustness checks, we first ensure that IV estimates do not suffer from endogeneity bias induced by firm location decisions, which could be influenced by the presence of infrastructures.¹⁵ To address an eventual location selectivity bias, we follow a similar approach to Dollar et al. (2006) and exclude from the sample large firms, exporters, and foreign firms, whose location decisions could be related to the presence of SMC landing stations or IXPs. In fact, various surveys point out that large or foreign firms are more geographically mobile, while smaller or domestic firms' location choices are mostly influenced by the location of their founders' social capital (Zaheer et al., 2009). Moreover, we also exclude exporters to make sure that estimated effects of firms' distances to SMC landing points do not reflect the effect of distance to ports (where SMCs can be laid) on their performance.

We start by conducting IV estimations on a sample of small and medium enterprises (SMEs) and report estimates in table 4. First-stage estimations support that despite this first sampling restriction, our instrument set is valid in sales, labour productivity, and employment equations (columns (1) to (3)). However, three-level IV estimates of the sales and productivity equations are found to be more than 1.5 times higher than IV-2SLS estimates in columns (1) to (3) and higher than previous IV estimates. They point to a larger local impact of Internet use on smaller firms' outcomes: a 10% increase in the local incidence of email use by firms raises their average sales by 38% and their average labour productivity by 31%, corresponding respectively to a 2.2 and a 1.6 standard deviation increase in these outcomes. The magnitude of these effects is confirmed by three-level IV estimations conducted on a sample this time restricted to domestic and non-exporter

¹⁵ This concern for the job equation is lowered by the recent research of Clarke et al. (2016), who stress that the presence of infrastructure is not robustly related to job growth.

SMEs, reported in table 5. They also point to very large local effects of email use on firms' sales and labour productivity equations. The effect of email use on temporary employment, however, turns insignificant with this sample restriction.¹⁶

Again, to make sure that country-specific effects do not bias coefficients, three-level dummy-variable IV estimations are conducted over the sample of domestic non-exporter SMEs and reported in Appendix C.4. These benchmark estimations in columns (1) and (2) strongly converge with the coefficient estimates of sales and productivity equations reported in tables 4 and 5 (columns (4) and (5)).

Table 4. Digital vulnerabilities and the performance of small and medium firms – IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	Total sales	Sales/ worker	Temp. emp.	Total sales	Sales/ worker	Temp. emp.
		IV-2SLS			3-level IV	
Email use	2.229** (0.937)	1.843** (0.864)	1.347* (0.800)	3.873*** (0.255)	3.118*** (0.196)	0.511*** (0.177)
First stage estimates:						
Firm distance to SMC/IXP	-0.015*** (0.005)	-0.015*** (0.005)	-0.014*** (0.005)	-0.0102*** (0.0027)	-0.0100*** (0.0033)	-0.0083*** (0.0026)
Centroid distance to SMC	-0.058*** (0.018)	-0.057*** (0.018)	-0.046*** (0.017)	0.0056 (0.011)	0.0058 (0.011)	0.0135 (0.0097)
Seaquake freq., 1000km rad.	-0.004* (0.002)	-0.004* (0.002)	-0.004* (0.002)	-0.004*** (0.0014)	-0.004*** (0.0012)	-0.0036*** (0.0013)
Hansen test (p. value)	0.96	0.47	0.13			
Weak ident. test (F-stat)	9.73***	9.70***	6.99***			
Under-ident. test (LM-stat)	23.74***	23.68***	18.06***			
				Random effect parameters		
Random intercepts				Country, sector	Country, sector	Country, sector
Random slope [email use]				Country, sector	Country, sector	Country, sector
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummies	Region, sector, year of interview, firm size					
<i>N</i>	2527	2524	2682	2527	2591	2484
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

¹⁶ Online Appendix A.1 and A.2 report first- and second-stage coefficients of three-level IV estimations separately conducted on a sample of non-exporters and domestic firms. They also stress a more-than-proportional, though softer, significant impact of email use on firms' revenue and labour productivity. A significant impact on temporary employment is found with the sample of non-exporters, but it turns insignificant when the sample is restricted to domestic firms.

Table 5. Digital vulnerabilities and the performance of domestic non-exporter SMEs – IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	Domestic non-exporter SMEs					
	IV 2SLS			3-level IV		
	Total sales	Sales/ worker	Temp. employment	Total sales	Sales/ worker	Temp. employment
Email use	1.082 (1.075)	1.168 (0.935)	1.568 (1.070)	4.152*** (0.274)	3.416*** (0.232)	0.140 (0.173)
1st stage estimates:						
Firm distance to SMC/IXP	-0.0135** (0.005)	-0.0135** (0.005)	-0.0113** (0.005)	-0.0101*** (0.0026)	-0.0100*** (0.0027)	-0.0086*** (0.0027)
Centroid distance to SMC	-0.059** (0.020)	-0.059** (0.020)	-0.044** (0.020)	0.0132 (0.0109)	-0.0132 (0.0115)	0.0180* (0.009)
Seaquake freq., 1000km rad.	-0.003 (0.002)	-0.003 (0.002)	-0.0025 (0.002)	-0.0036*** (0.0012)	-0.0036*** (0.0012)	-0.0033** (0.0014)
Hansen test (p. value)	0.93	0.80	0.17			
Weak ident. test (F-stat)	6.81***	6.80***	4.81***			
Under-ident. test (LM-stat)	17.42***	17.40***	13.09***			
Random intercepts				Country, sector	Country, sector	Country, sector
Random slope [email use]				Country, sector	Country, sector	Country, sector
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummies	Region, sector, year of interview, firm size					
<i>N</i>	2574	2571	2559	2344	2341	2470
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

As a last check for location bias, we run the same estimations on a sample restricted to firms created before the year the cable was activated (table 6). In fact, location bias is unlikely for firms that have chosen their location before SMC arrival. Our results stress the positive and significant effect of email use on firm sales, productivity, and employment. The magnitude of the effect of email use on sales and productivity lies within the same range as previous estimates in tables 4 and 5.

Table 6. Digital vulnerabilities and the performance of firms created before SMC's arrival – IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	IV 2SLS			3-level estimations		
	Total sales	Sales/ worker	Temp. employment	Total sales	Sales/ worker	Temp. employment
Email use	2.202* (1.168)	2.605** (1.091)	1.794 (1.329)	3.933*** (0.240)	3.124*** (0.291)	0.493** (0.243)
1st stage estimates:						
Firm distance to SMC/IXP	-0.012*** (0.005)	-0.012*** (0.005)	-0.009** (0.004)	-0.0067** (0.00294)	-0.0068** (0.0029)	-0.0057** (0.00283)
Centroid distance to SMC	-0.062*** (0.021)	-0.063*** (0.021)	-0.053*** (0.019)	-0.0031 (0.0119)	-0.0031 (0.0095)	0.0038 (0.0096)
Seaquake freq., 1000km rad.	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)	-0.0040*** (0.0011)	-0.0040*** (0.0013)	-0.0036*** (0.0012)
Hansen test (p. value)	0.86	0.40	0.62			
Weak ident. test (F-stat)	6.18***	6.17***	4.73***			
Under-ident. test (LM-stat)	16.32***	16.30***	12.67***			
Random intercepts				Country, sector	Country, sector	Country, sector
Random slope [email use]				Country, sector	Country, sector	Country, sector
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummies	Region, sector, year of interview, firm size					
<i>N</i>	2431	2428	2559	2431	2428	2561
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

5.2. Other robustness checks

Next, we estimate the effect of fast Internet on firm outcomes using alternative proxies of firms' Internet access: a dummy variable equal to 1 if the firm communicates with its clients and suppliers via website. Results, reported in columns (1) to (3) of table 6, support a positive and significant effect of website use by firms on their sales, productivity, and employment level. These estimates are of similar magnitude to estimates of the effect of email use on firms' outcomes.

Table 7. Website use, digital vulnerabilities, and firm performance – IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IV-2SLS				3-level IV		
	Total sales	Sales/ worker	Employment		Total sales	Sales/ worker	Employment
Website use	2.639** (1.149)	2.607** (1.051)	2.721* (1.472)	2.585** (1.286)	3.185*** (0.342)	2.734*** (0.299)	-0.113 (0.237)
1st stage estimates:							
Firm distance to SMC/IXP	-0.014*** (0.005)	-0.014*** (0.005)	-0.013*** (0.005)	0.014*** (0.005)	-0.011*** (0.002)	-0.011*** (0.002)	-0.010*** (0.002)
Centroid distance to SMC	-0.051*** (0.023)	-0.051*** (0.023)	-0.024 (0.020)	-	-0.0138 (0.017)	-0.0140 (0.017)	-0.0003 (0.018)
Seaquake freq., 1000km rad.	-0.002 (0.002)	-0.002 (0.002)	-0.0016 (0.0017)	-	-0.003*** (0.001)	-0.003*** (0.001)	-0.0026** (0.001)
Hansen test (p. value)	0.76	0.50	0.28	-			
Weak ident. test (F-stat)	5.27***	5.27***	3.12**	8.55***			
Under-ident. test (LM-stat)	13.92***	13.92***	9.04**	8.06***			
Random intercepts					Country, sector	Country, sector	Country, sector
Random slope [website use]					Country, sector	Country, sector	Country, sector
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummies	Region, sector, year of interview, firm size						
<i>N</i>	2631	2628	2795	2795	2631	2632	2798
<i>Countries</i>	62	62	66	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

As a last robustness check, because the World Bank Enterprise Surveys also consider the firm’s size category (small, medium, or large) as sample stratification criteria, we conduct four-level IV estimation by adding this level of analysis in data aggregation¹⁷ and in the hierarchical modelling of the data. In other words, using data averaged by firm size, we re-estimate equations (2’) and (3’), applying a four-level random-slope model, that is, keeping the random coefficient at the country level but making the random intercept $\Gamma_0 = \gamma'_0 + \gamma''_{0,j} + \gamma'''_{0,s,j} + \gamma''''_{0,s,j,k}$ with k representing either small, medium, or large firm categories.

¹⁷ Firm-level data of large, medium, and small companies from the same city, same sector, same interview year, and same country are therefore averaged separately.

Table 8. 4-level IV estimations

	(1)	(2)	(3)	(5)	(6)	(7)
	Sales	Sales/worker	Employment	Sales	Sales/worker	Employment
<i>Email use</i>	2.600** (1.019)	2.146** (0.957)	1.827** (0.899)	3.759*** (1.318)	2.927** (1.149)	-0.151 (1.049)
	IV 2SLS			4-level IV		
1st stage estimates:						
Firm distance to SMC/IXP	-0.013*** (0.004)	-0.013*** (0.004)	-0.0096** (0.004)	-0.0111** (0.00432)	-0.0110** (0.0043)	-0.00969** (0.0040)
Centroid distance to SMC	-0.052*** (0.0159)	-0.052*** (0.0159)	-0.045*** (0.015)	0.0101 (0.0200)	0.010 (0.020)	0.0125 (0.019)
Seaquake freq., 1000km rad.	-0.005*** (0.002)	-0.005*** (0.002)	-0.005*** (0.002)	-0.0037*** (0.00118)	-0.0036*** (0.0012)	-0.00304** (0.0012)
Hansen test (p. value)	0.62	0.33	0.39			
Weak ident. test (F-stat)	11.60***	11.57***	8.61***			
Under-ident. test (LM-stat)	31.22***	31.15***	23.55***			
Random intercepts				Country, sector, location	Country, sector, location	Country, sector, location
Random slope [email use]				Country	Country	Country
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
Dummies	Region, sector, year of interview, firm size					
<i>N</i>	5442	5442	5851	5438	5433	5843
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations.

6- Discussion and concluding remarks

This paper provides evidence of a large and positive impact of NICT on firms' performance at the local level. An increased use of the Internet (either email or website use) is found to be a consistent way to boost sales, productivity, and in a less consistent way, temporary employment. The effect of Internet use on firms' revenue and labour productivity gets even larger when we focus the analysis on smaller or inward-oriented firms, thereby corroborating the finding of Clarke et al. (2015) of a stronger effect of Internet use on small businesses.

Through the IV approach followed in this paper, we also pointed out the contribution of maritime and terrestrial fast Internet infrastructure to the economy through the prism of the concomitant digital vulnerabilities some developing and transition countries may be subject to: In fact, in some countries, the arrival of fibre has boosted the Internet economy as a whole but may have also widened the digital divide between coastal and inland, remote and geographically integrated, or rural and urban areas and increased the economy's exposure to fibre cable breaks and Internet shutdowns.

According to multilevel IV estimations, a 10% increase in the use of email in the population of firms raises by 26% the firms' average annual sales, by 21% average sales per worker, and by 6% the average number of temporary workers, though this last estimate appears to be less robust across specifications. First, exploiting an original database on telecommunications infrastructures' deployment worldwide, we find that firms or countries that are remote from IXPs or SMC landing stations, or countries with large territories, face geographical handicaps that may deprive them of the potential benefits of the digital revolution (Röller & Waverman, 2001). Second, we show that even if digital isolation in various developing countries has been significantly reduced by the laying of hundreds of SMCs across the world, many countries remain highly vulnerable to cable breaks, whether they are caused by seaquakes, piracy, fishing, or anchoring. Our results specifically stress the importance of a country's exposure to seaquake-induced SMC breaks for the development of the Internet economy and the performance of firms, but the underlying mechanisms can be extended to other previously mentioned sources of cable faults. Last, our results stress that estimated impact of Internet access on firm outcomes get stronger for domestic firms, SMEs, and non-exporters.

All in all, IV estimations stress that digital vulnerabilities induced by the worldwide deployment of wireline telecommunications infrastructures impacts developing economies by creating macro-, meso- and micro-level heterogeneities in terms of Internet access that probably lower the potential benefits of the digital economy in the developing world.

References

- Bates, P. (2014, April). Submarine cables in Sub-Saharan Africa: terrestrial networks need to keep up. London: Analysys Mason.
- Borland, J. (2008, February 5). Analyzing the Internet collapse: Multiple fiber cuts to undersea cables show the fragility of the Internet at its choke points. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/s/409491/analyzing-the-Internet-collapse/>.
- Carter, L., Burnett, D., Drew, S., Marle, G., Hagadorn, L., Bartlett-McNeil, D., & Irvine, N. (2009). *Submarine cables and the oceans – Connecting the world* (UNEP-WCMC Biodiversity Series No. 31). Cambridge: UNEP-WCMC.
- Center for International Earth Science Information Network. (2015). *Gridded population of the world, version 4 (GPWv4): Population count*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4BG2KXS>.
- Cette, G., Clerc, C., & L. Bresson. (2015). Contribution of ICT diffusion to labour productivity growth: The United-States, Canada, the Eurozone, and the United-Kingdom, 1970–2013. *International Productivity Monitor*, 28: 81–88.
- Choi, C. (2010). The effect of the Internet on service trade. *Economics Letters*, 109(2): 102–104.
- Choi, C., & Yi, M. H. (2009). The effect of the Internet on economic growth: Evidence from cross-country panel data. *Economics Letters*, 105(1): 39–41.
- Chowdhury, S. K., & Wolf, S. (2003). Use of ICTs and the economic performance of SMEs in East Africa. WIDER Discussion Paper no. 2003/06: World Institute for Development Economics.
- Clark, B. (2016). Undersea cables and the future of submarine competition. *Bulletin of the Atomic Scientists*, 72(4): 234–237.
- Clarke, G. (2008). Has the Internet increased exports for firms from low and middle-income countries? *Information Economics and Policy*, 20(1): 16–37.
- Clarke, G., & Wallsten, S. J. (2006). Has the Internet increased trade? Developed and developing country evidence. *Economic Inquiry*, 44(3): 465–84.
- Clarke, G., Qiang, C. Z., & Xu, L. C. (2015). The Internet as a general purpose technology: Firm-level evidence from around the world. *Economics Letters*, 135(C): 24–27.
- Clarke, G., Li, Y., & Xu, L. C. (2016). Business environment, economic agglomeration and job creation around the world. *Applied Economics*, 48(33), 3088–3103.
- Czernich, N., Falck, O., Kretschner, T., & Woessmann, L. (2011). Broadband infrastructure and economic growth. *The Economic Journal*, 121(552): 205–532.
- Dollar, D., Hallward-Driemeier, M., & Mengistae, T. (2006). Investment climate and international integration. *World Development*, 34(9): 1498–1516.
- Freund, C., & Weinhold, D. (2002). The Internet and international trade in services. *American Economic Review*, 92(2), 236–240.
- Freund, C., & Weinhold, D. (2004). The Effect of the Internet on international trade. *Journal of International Economics*, 62(1): 171–189.
- Head, K., Ries, J., & Swenson, D. (1995). Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States. *Journal of International Economics*, 38(3–4), 223–247.
- Hjort, J., & Poulsen, J. (2017). *The arrival of fast Internet and employment in Africa* (Working Paper No. w23582). Cambridge, MA: National Bureau of Economic Research.

- Hox, J.J. (2010). *Multi-level analysis, techniques and applications*. New York: Routledge, 2010.
- Litan, R. E., & Rivlin, A. M. (2001). Projecting the economic impact of the Internet. *American Economic Review*, 91(2), 313-317.
- Moulton, B.R. (1990). An illustration of pitfall in estimating the effects of aggregate variables on micro-units. *Review of Economics and Statistics*, 72(2): 334-338.
- Organisation for Economic Co-operation and Development (2014). International cables, gateways, backhaul and international exchange points. Working Party on Communication Infrastructures and Services Policy, Directorate for Science, Technology and Industry Committee for Information, Computer and Communications Policy.
- Palmer-Felgate, A., Irvine, N., Ratcliffe, S., & Bah, S. S. (2013, April 22-25). *Marine maintenance in the zones – A global comparison of repair commencement times*. Paper presented at Suboptic Conference: From Ocean to Cloud, Paris.
- Paunov, C., & Rollo, V. (2015). Overcoming obstacles: The Internet's contributions to firm development. *World Bank Economic Review*, 29(suppl.1): S192-S204.
- Röller, L-H, & Waverman, L. (2001). Telecommunications infrastructure and economic development: A simultaneous approach. *American Economic Review*, 91(4): 909-923.
- Schumann, R., & Kende, M. (2013, May). Lifting barriers to Internet development in Africa: Suggestions for improving connectivity (Report for the Internet Society). London: Analysys Mason.
- Soh, W., Machiyama, H., Shirasaki, Y., & Kasahara, J. (2004). Deep-sea floor instability as a cause of deepwater cable fault, off eastern part of Taiwan. *Frontier Research of Earth Evolution*, 2: 1-8.
- Stoker T.M. (2010) Aggregation (econometrics). In: Durlauf S.N., Blume L.E. (eds) *Macroeconometrics and Time Series Analysis*. The New Palgrave Economics Collection. Palgrave Macmillan, London.
- Tian, L. (2018). *Division of labor and extent of the market: Theory and evidence from Brazil* (Job Market Paper). Retrieved from <https://lin-tian.github.io/files/JobMarketPaper.pdf>.
- Towela, N-J., & Tesfaye, B. (2015, May 22). *Internet development and Internet governance in Africa*. Internet Society. Retrieved from <https://www.Internetsociety.org/resources/doc/2015/Internet-development-and-Internet-governance-in-africa/>.
- TeleGeography. (2016). *Submarine cable map*. Available at <http://www.submarinecablemap.com/>.
- Wallsten, S. (2005). Regulation and internet use in developing countries. *Economic Development and Cultural Change*, 53(2), 501-523.
- Weller, D., & Woodcock, B. (2013, January). *Internet traffic exchange: Market developments and policy challenges* (OECD Digital Economy Papers, No. 207). Paris: OECD Publishing.
- Widmer, G., Wolff, M. (Producers), & Dotzer, U. (Director). (2010). *Les câbles Internet sous-marins: L'autoroute mondiale de l'information* [motion picture]. Arte reportage, Fact+ film production.
- Zaheer, S., Lamin, A., & Subramani, M. (2009). Cluster capabilities or ethnic ties? Location choice by foreign and domestic entrants in the services offshoring industry in India. *Journal of International Business Studies*, 40(6): 944.

Appendix A. Data collection and treatment

A.1. Infrastructure deployment variables

Raw data on SMCs are drawn from Telegeography:

- All cables with date of commissioning
- All the landing stations of cables and their GPS coordinates
- The number and identity of telecoms operator-owners of cables

Raw data on Internet Exchange Points are drawn from Telegeography and completed by the *Packet Clearing House* and *Peering DB* databases:

- All IXPs with their status (active/inactive/project)
- their year of activation
- their GPS coordinates

After a conversion into polygons (disk with 5km diameter) to avoid topological inaccuracies, the SMC landing points and IXPs from each country are identified, located, and counted. Then, for each country, all cables related to these points and all IXPs are identified, which gives **the number of cables** and **the number of IXPs** variables. Finally, the set of the foreign landing points linked to these cables are identified, enabling the calculation of **the connectedness variable**, which counts the number of partner countries directly connected to the country of interest.

The age of the first cable is obtained by calculating the difference between the current year and the year of the first SMC's activation for each country. This variable is forward-looking and can take negative values at time t when the activation year occurs at time $t + k$.

Using information from Telegeography on SMC ownership structure, **the number of cable owners** is calculated for each country by summing the number of cable owners associated with all SMCs laid in that country.

A.2. Infrastructure gap proxies

We compute two infrastructure gap proxies: the structural need for infrastructure deployment and firms' need for infrastructure deployment. A distance raster map is defined from all coordinate points, which gives the distance of each firm from the nearest Internet Exchange Point.

The structural need for infrastructure deployment

Statistical inputs: SMC landing station coordinates, countries' centroids, spatial distribution of the population.

a. Country with cables: From the SMC landing points of a given country, the distance to each point of its territory is calculated in the form of a raster map with the Spatial Analyst's Cost Distance tool, using the Winkel III projection. The Zonal Statistics tool then gives us the distance from centroid of the country to the closest SMC landing station.

b. Country without cables: From the closest foreign SMC landing points, the distance to each terrestrial point of the world is calculated as previously.

The firm need for infrastructure deployment

Statistical inputs: SMC landing station coordinates, IXP coordinates, firms' city location coordinates.

Firms' city location coordinates (WBES), SMC landing station coordinates (Telegeography), and IXP coordinates (Telegeography, Peering DB) give points for which the distances to SMCs are calculated using the previously calculated distance raster. In some countries, firms may have been pooled and interviewed by province rather than city of location. In this case, we take the province's centroid as the firm's location coordinates. The firm infrastructure need is the minimum distance for the firm to reach the closest infrastructure node: either a landing station or an IXP.

A.3. Exposure to seaquake-induced cable faults

The Northern California Earthquake Data Center of the University of California, Berkeley, provides a global database of earthquakes. For each country, we get for each year the number, the location, and the average magnitude of epicentres of occurring seaquakes and are therefore able to compute the annual frequency of seaquakes within a 1000km radius of the stations.

To ensure that we do not take into account seaquakes that could induce tsunamis, which would hence violate restriction identification conditions, we do not take into account seaquakes whose magnitudes exceed 6.5 on the Richter scale. To ensure that we do take into account seaquakes that are strong enough to induce cable faults, we only count seaquakes with magnitudes exceeding 5 on the Richter scale. All in all, seaquakes considered for the empirical analysis are those occurring within a 1000km radius from SMC landing stations, whose magnitudes are between 5 and 6.5 on the Richter scale.

Appendix B. Summary statistics

B.1. Sample: 66 countries, 32,543 firm-level observations

	min	mean	max	std. dev	Sources
Dependent variables					
Ln total sales (USD)	0	13.28573	27.29597	2.824788	
Ln sales per worker (USD)	0	10.1028	23.51205	2.312553	WBES
Ln # fulltime temporary workers	0	0.9409591	10.12671	1.395419	
Independent variables					
Dummy email use	0	0.7012825	1	0.457702	WBES
Dummy website use	0	0.4469358	1	0.497183	
# Cables	0	1.830901	7	2.162687	
# IXPs	0	1.095315	9	1.611337	
Connectedness	0	10.42383	45	12.90558	
First cable seniority	-3	5.787405	21	6.514196	
# operators by country	0	12.23985	46	14.04888	Authors
Ln firm distance to SMCs or IXPs	0	3.551599	7.393878	2.592437	
Ln centroid distance to SMC landing stations	2.888737	6.150347	7.805853	1.043988	
Seaquake freq. 1000km rad from SMC	0	6.705904	40	12.31735	Authors
Landlockedness	0.3262622	1	0	0.4688513	
Ln GDP per cap	5.258385	7.310903	9.832184	1.117845	World Bank
Ln population	16.82063	21.02389	13.46996	1.608682	
Gross enrolment rate 1ary	105.3382	149.9517	53.10298	15.06023	
Polity2	2.957631	10	-7	6.041872	Polity IV
Ln firm age	0	2.642401	4.744932	0.7778525	
City size	1	2.583894	5	1.260414	
% of working capital financed by external sources	0	11.52634	100	22.74581	
Dummy state-owned	0	0.0172249	1	0.130110	WBES
Dummy foreign-owned	0	0.1016734	1	0.3022226	
Dummy large size	0	0.1791278	1	0.3834648	
% of direct and indirect exports	0	9.72468	100	24.68762	
Electricity access obstacle	0	1.694493	4	1.477965	

B.2. Summary statistics, location-level data, unrestricted sample

	Observations (countries)	min	mean	max	std. dev
Dependent variables					
Ln total sales (USD)	2637(62)	5.047455	13.04806	23.65241	2.335392
Ln sales per worker (USD)	2634(62)	3.012011	9.814647	18.95393	1.929685
Ln # fulltime temporary workers	2800(66)	0	1.023374	6.671392	0.9477663
Variables of interest					
Dummy email use	2800(66)	0	0.6699535	1	0.3415791
Dummy website use	2800(66)	0	0.4189243	1	0.3391122

B.3. Summary statistics, location-level data, sample of SMEs

	Observations (countries)	min	mean	max	std. dev
Dependent variables					
Ln total sales (USD)	2527(62)	5.047455	12.52796	21.81643	2.142998
Ln sales per worker (USD)	2524(62)	3.012011	9.745639	18.94806	1.929685
Ln # fulltime temporary workers	2682(66)	0	0.8582599	4.875197	0.8359684
Independent variables					
Dummy email use	2682(66)	0	0.6495465	1	0.3556639
Dummy website use	2682(66)	0	0.3857633	1	0.347084

B.4. Sample composition

IV sample: 66 countries, 2,787 city-level observations, 32,543 firm-level observations

iso	2007	2009	2010	2011	2012	2013	2014	Total	iso	2007	2009	2010	2011	2012	2013	2014	Total	
AFG	0	0	0	0	0	0	51	51	LSO	0	11	0	0	0	0	0	11	
ALB	36	0	0	0	0	0	0	36	MDA	0	0	0	0	0	36	0	36	
ARG	0	0	55	0	0	0	0	55	MDG	0	32	0	0	0	58	0	90	
AZE	0	0	0	0	0	35	0	35	MLI	31	0	30	0	0	0	0	61	
BEN	0	18	0	0	0	0	0	18	MNG	0	0	0	0	0	31	0	31	
BFA	0	21	0	0	0	0	0	21	MOZ	35	0	0	0	0	0	0	35	
BGR	35	0	0	0	0	0	0	35	MUS	0	49	0	0	0	0	0	49	
BLR	0	0	0	0	0	18	0	18	MWI	0	31	0	0	0	0	0	31	
BOL	0	0	33	0	0	0	0	33	NER	0	21	0	0	0	0	0	21	
BTN	0	34	0	0	0	0	0	34	NIC	0	0	22	0	0	0	0	22	
CAF	0	0	0	3	0	0	0	3	NPL	0	31	0	0	0	31	0	62	
CHN	0	0	0	0	336	0	0	336	PAN	0	0	21	0	0	0	0	21	
CIV	0	21	0	0	0	0	0	21	PER	0	0	40	0	0	0	0	40	
CMR	0	29	0	0	0	0	0	29	PRY	0	0	22	0	0	0	0	22	
COG	0	16	0	0	0	0	0	16	RWA	0	0	0	4	0	0	0	4	
CRI	0	0	21	0	0	0	0	21	SEN	37	0	0	0	0	0	0	37	
DJI	0	0	0	0	0	11	0	11	SLV	0	0	22	0	0	0	0	22	
DOM	0	0	23	0	0	0	0	23	SRB	0	0	0	0	0	35	0	35	
ECU	0	0	32	0	0	0	0	32	SVK	0	0	0	0	0	35	0	35	
ERI	0	23	0	0	0	0	0	23	SVN	0	0	0	0	0	25	0	25	
EST	0	0	0	0	0	50	0	50	TCD	0	12	0	0	0	0	0	12	
ETH	0	0	0	50	0	0	0	50	TGO	0	12	0	0	0	0	0	12	
GEO	0	0	0	0	0	45	0	45	TJK	0	0	0	0	0	33	0	33	
GHA	36	0	0	0	0	47	0	83	TZA	0	0	0	0	0	55	0	55	
GTM	0	0	22	0	0	0	0	22	UGA	0	0	0	0	0	64	0	64	
GUY	0	0	2	0	0	0	0	2	UKR	0	0	0	0	0	35	0	35	
HND	0	0	33	0	0	0	0	33	URY	0	0	22	0	0	0	0	22	
KAZ	0	0	0	0	0	57	0	57	VEN	0	0	29	0	0	0	0	29	
KGZ	0	0	0	0	0	34	0	34	VNM	0	65	0	0	0	0	0	65	
LAO	0	43	0	0	46	0	0	89	YEM	0	0	55	0	0	24	0	79	
LBN	0	0	0	0	0	78	0	78	ZAF	45	0	0	0	0	0	0	45	
LBR	0	21	0	0	0	0	0	21	ZAR	0	0	32	0	0	43	0	75	
LKA	0	0	0	73	0	0	0	73	ZMB	38	0	0	0	0	45	0	83	
										Total	293	490	516	130	382	925	51	2,787

Appendix C. Benchmark estimations

C.1. Three-level first-stage IV estimates – unrestricted sample

	(1)	(2)	(3)
	Total sales	Sales/worker	Temp. employment
	Instrumented Variable: Email use		
Firm distance to SMC/IXP	-0.00805*** (0.00232)	-0.00793*** (0.00227)	-0.00702*** (0.00267)
Centroid distance to SMC	0.00405 (0.0109)	0.00424 (0.00923)	0.00917 (0.00805)
Seaquake freq., 1000km rad.	-0.00350*** (0.00123)	-0.00344*** (0.0009)	-0.00333*** (0.0012)
# SMCs	0.0161 (0.0108)	0.0156* (0.00914)	0.0136 (0.00881)
# IXPs	0.00995* (0.00558)	0.0101* (0.00514)	0.0213*** (0.00459)
Time of 1st cable	-0.00417 (0.00260)	-0.00395* (0.00228)	-0.00249 (0.00247)
Landlocked	0.0383* (0.0198)	0.0383** (0.0179)	0.0427** (0.0168)
GDP per cap.	0.0682*** (0.00846)	0.0681*** (0.00902)	0.0741*** (0.00947)
Population	0.00706 (0.0105)	0.00705 (0.0127)	-0.00545 (0.00920)
Education	0.00140*** (0.000465)	0.00141*** (0.000488)	0.00158*** (0.000537)
Democracy	0.00744*** (0.00165)	0.00745*** (0.00174)	0.00650*** (0.00153)
State-owned	0.0139 (0.0559)	0.0144 (0.0709)	0.00199 (0.0577)
Foreign	0.143*** (0.0355)	0.143*** (0.0311)	0.102*** (0.0345)
Age	0.0179 (0.0158)	0.0164 (0.0186)	0.0241* (0.0127)
City size	-0.0426*** (0.00525)	-0.0427*** (0.00537)	-0.0449*** (0.00481)
External funding	0.000780 (0.000568)	0.000762 (0.000464)	0.00104** (0.000475)
Electricity obstacle	0.0159** (0.00789)	0.0154 (0.00965)	0.0185*** (0.00588)
% of exports	0.000805** (0.000401)	0.000800** (0.000359)	0.000707 (0.000435)
Dummies	Region, year, sector, firm size		
Random effect parameters			
Random intercept country	-2.455*** (0.0850)	-2.456*** (0.106)	-2.541*** (0.0911)
Random intercept sector	-2.710*** (0.0519)	-2.706*** (0.0585)	-2.769*** (0.0535)
<i>N</i>	2633	2630	2797
<i>Countries</i>	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses and are robust to heteroscedasticity. First-stage three-level estimates result from a 100 replications bootstrap.

C.2. Country-dummy variable OLS and three-level estimates of equation (1)

	(1)		(2)		(3)		(4)		(5)		(6)	
	Total sales		Sales/worker		Temp employment							
	OLS	3-level	OLS	3-level	OLS	3-level	OLS	3-level	OLS	3-level	OLS	3-level
Email use	0.735***	0.820***	0.671***	0.661***	-0.0723	0.248***	(0.164)	(0.123)	(0.133)	(0.122)	(0.138)	(0.0853)
State-owned	-0.240	0.362	-0.253	0.316	-0.0496	0.214	(0.289)	(0.222)	(0.258)	(0.242)	(0.205)	(0.257)
Foreign	-0.0779	0.626***	0.168	0.427***	0.0383	0.0778	(0.349)	(0.193)	(0.215)	(0.161)	(0.287)	(0.137)
Age	0.020	0.224***	-0.0628	0.118	-0.0281	-0.0105	(0.117)	(0.0766)	(0.0940)	(0.0736)	(0.101)	(0.0410)
City size	-0.158***	0.0404	-0.111**	0.0478	-0.0496	0.0201	(0.046)	(0.0838)	(0.0438)	(0.0826)	(0.0403)	(0.0233)
External funding	0.0060	0.00662***	0.00903***	0.00487**	0.00311	0.00528***	(0.0041)	(0.00192)	(0.00248)	(0.00206)	(0.00308)	(0.00120)
Electricity obstacle	-0.0958	0.00389	-0.0929	0.00624	0.0391	0.0395**	(0.074)	(0.0386)	(0.0578)	(0.0384)	(0.0534)	(0.0184)
% of exports	-0.0017	0.00411**	-0.00468	0.00198	0.00300	0.000548	(0.0028)	(0.00190)	(0.00301)	(0.00179)	(0.00409)	(0.00144)
Dummies	Region, country, year, sector, firm size											
N	3,372	3,372	3369	3369	3560	3560						
Countries	67	67	67	67	71	71						
R2	0.76	-	0.62	-	0.33	-						

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses and are robust to heteroscedasticity.

C.3. Country-dummies three-level IV estimations – unrestricted sample

	(4)	(5)	(6)
	Total sales	Sales/worker	Temp. employment
Email use	4.534**	4.045**	0.441
	(1.665)	(1.591)	(1.374)
Micro controls	Yes	Yes	Yes
Dummies	Region, country, year, sector		
Random effect parameters			
Random intercepts	Country, sector		
Random slopes [email use]	Country, sector		
1st stage estimates			
Firm distance to SMC/IXP	-0.0088*	-0.0086*	-0.0079*
	(0.0047)	(0.0047)	(0.0044)
Micro controls	Yes	Yes	Yes
Dummies	Country, year, sector, firm size		
Random intercepts:	Country, sector		
N	2637	2634	2800
Countries	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses and are robust to heteroscedasticity. Control estimates not reported.

C.4. Country-dummies three-level IV estimations – domestic non-exporter SMEs

	(1)	(2)	(3)
	Total sales	Sales/worker	Temp. employment
Email use	3.826*** (1.439)	3.279*** (1.198)	0.400 (0.732)
Micro controls Dummies	Yes Region, country, year, sector, firm size	Yes	Yes
Random effect parameters			
Random intercepts	Country, sector		
Random slopes [email use]	Country		
1st stage estimates			
Firm distance to SMC/IXP	-0.0114** (0.0051)	-0.0114** (0.0051)	-0.0103** (0.0046)
Micro controls Dummies	Yes Country, year, sector, medium-size firm	Yes	Yes
Random intercepts:	Country, sector		
<i>N</i>	2344	2341	2470
<i>Countries</i>	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors are presented in parentheses and are robust to heteroscedasticity. Control estimates not reported.

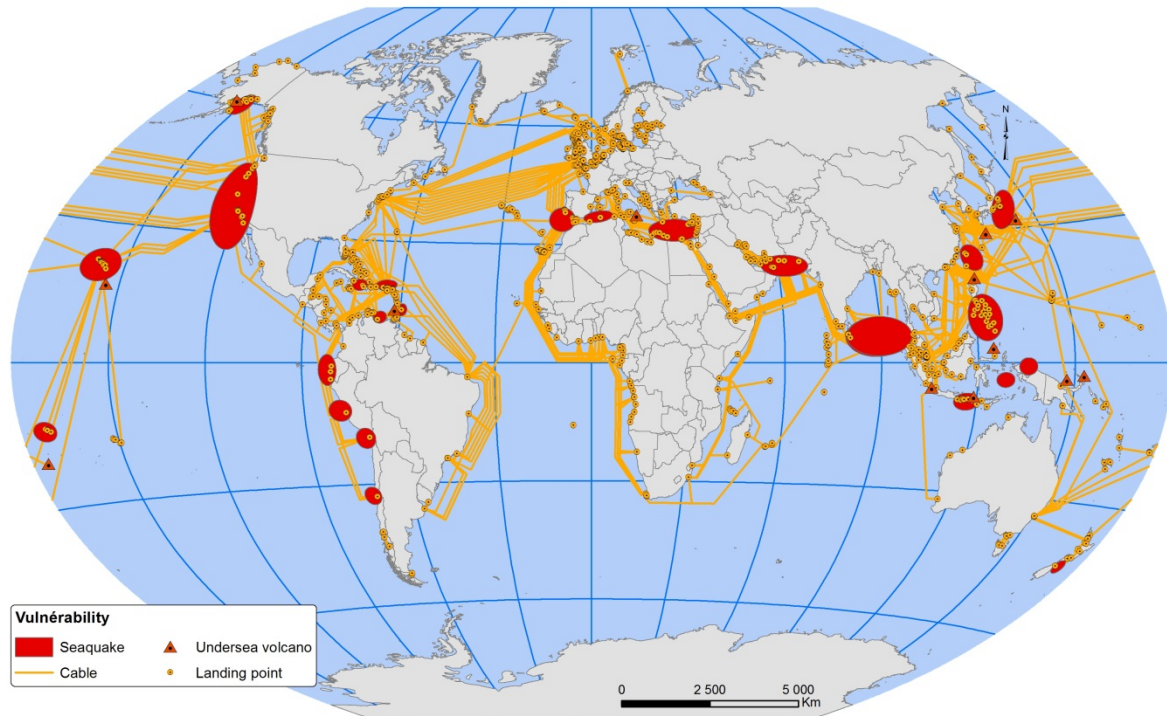
Appendix D. Infrastructure deployment, centroid, and firm location in China



Sources: Authors. Raw data: Telegeography, Packet Clearing House and Peering DB databases.

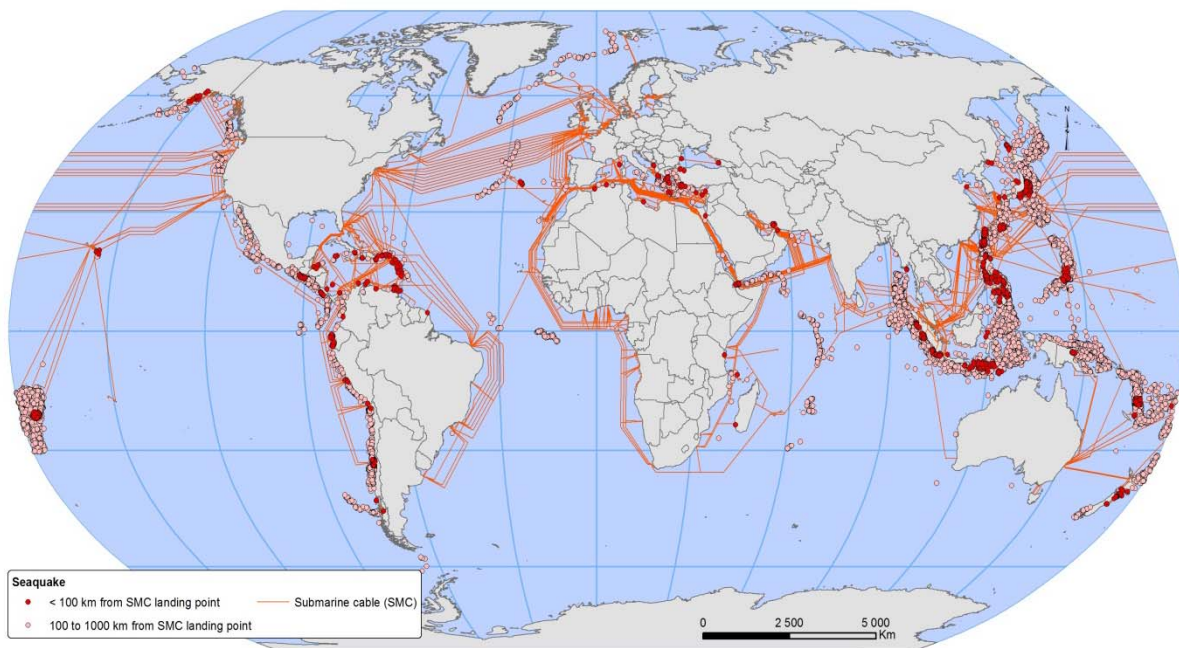
Appendix E. Digital vulnerability to seismic risk

Map E.1. Identified areas of digital vulnerability to seismic and volcanic risk



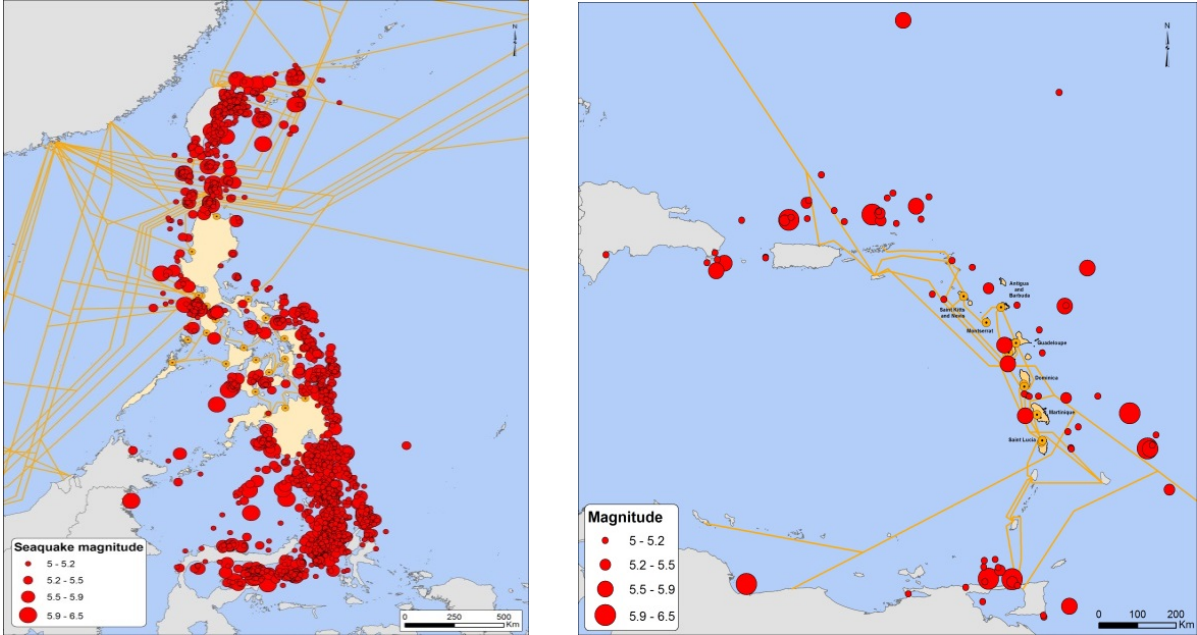
Sources: Authors. Raw data on SMC is drawn from Telegeography. Vulnerability areas identified by Dr Raphaël Paris, Research Officer in volcanology at CNRS and Laboratoire Magmas et Volcans (LMV), Observatoire de Physique du Globe de Clermont-Ferrand, Clermont-Auvergne University. This map is not exhaustive but based on own his knowledge of volcanic and seismic hazards and on events documented in catalogues and databases.

Map E.2. International seismic activity around SMC landing stations, 2005–2015



Sources: Authors. Raw data: Telegeography and Northern California Earthquake Data Center of the University of California, Berkeley.

Map E.3. Sequake occurrence within a 1000km radius from SMC landing stations in Philippines and the West Indies, 2005–2015



Sources: Authors. Raw data: Telegeography and Northern California Earthquake Data Center of the University of California, Berkeley.

Map E.4. SMC maintenance areas



Source: Global Marine Systems Ltd.

Online Appendices

Digital vulnerabilities and the performance of non-exporters – IV estimations

	(7)	(8)	(9)	(10)	(11)	(12)
	Total sales	Sales/ worker	Temp. emp.	Total sales	Sales/ worker	Temp. emp.
	IV-2SLS			3-level IV		
Email use	1.676*	1.611*	1.786**	2.851***	2.268***	0.360**
	(0.972)	(0.869)	(0.830)	(0.252)	(0.237)	(0.156)
First stage estimates:						
Firm distance to SMC/IXP	-0.015***	-0.015***	-0.013**	-0.015***	-0.015***	-0.013**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Centroid distance to SMC	-0.057**	-0.057**	-0.041**	-0.057**	-0.057**	-0.041**
	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)
Seaquake freq., 1000km rad.	-0.003*	-0.003*	-0.002 (0.002)	-0.003*	-0.003*	-0.002
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Random intercepts				Country, sector	Country, sector	Country, sector
Random slope [email use]				Country, sector	Country, sector	Country, sector
Hansen test (p. value)	0.96	0.47	0.13			
Weak ident. test (F-stat)	9.73***	9.70***	6.99***			
Under-ident. test (LM-stat)	23.74***	23.68***	18.06***			
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	2514	2511	2637	2461	2458	2584
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

Digital vulnerabilities and the performance of domestic firms – IV estimations

	(1)	(2)	(3)	(4)	(5)	(6)
	Total sales	Sales/ worker	Temp. emp.	Total sales	Sales/ worker	Temp. emp.
	IV 2SLS			Bootstrapped IV 3-level		
	Total sales	Sales/ worker	Temp. employment	Total sales	Sales/ worker	Temp. employment
Email use	2.169** (1.030)	2.183** (0.960)	2.147 (1.578)	2.676*** (0.245)	2.140*** (0.230)	0.0805 (0.196)
1st stage estimates:						
Firm distance to SMC/IXP	-0.014*** (0.005)	-0.014*** (0.005)	-0.011** (0.005)	-0.0076*** (0.00264)	-0.0075*** (0.00290)	-0.0064** (0.00273)
Centroid distance to SMC	-0.061** (0.019)	-0.061** (0.019)	-0.043** (0.020)	0.0097 (0.0105)	0.00975 (0.0112)	0.0104 (0.0095)
Seaquake freq., 1000km rad.	-0.003* (0.002)	-0.003* (0.002)	-0.003 (0.002)	-0.0035*** (0.0012)	-0.0034** (0.0014)	-0.0033** (0.0013)
Hansen test (p. value)	0.84	0.41	0.07			
Weak ident. test (F-stat)	9.38***	9.38***	5.91***			
Under-ident. test (LM-stat)	23.45***	23.37***	16.09***			
Random intercepts				Country, sector	Country, sector	Country, sector
Random slope [email use]				Country, sector	Country, sector	Country, sector
BBI, macro, micro controls	Yes	Yes	Yes	Yes	Yes	Yes
<i>N</i>	2428	2425	2559	2574	2571	2559
<i>Countries</i>	62	62	66	62	62	66

Note: * significant at 10%, ** significant at 5%, *** significant at 1%. Control estimates not reported. Standard errors are presented in parentheses and are robust to heteroscedasticity. Standard errors are clustered by country-year-sector in IV-2SLS estimations. Three-level second-stage estimates result from a 50 replications double-bootstrap. Reported first-stage three-level estimates result from a 100 replications bootstrap.

“Sur quoi la fondera-t-il l'économie du monde qu'il veut gouverner? Sera-ce sur le caprice de chaque particulier? Quelle confusion! Sera-ce sur la justice? Il l'ignore.”

Pascal



Created in 2003 , the **Fondation pour les études et recherches sur le développement international** aims to promote a fuller understanding of international economic development and the factors that influence it.

 **Contact**

www.ferdi.fr

contact@ferdi.fr

+33 (0)4 73 17 75 30