



European
University
Institute

ROBERT
SCHUMAN
CENTRE FOR
ADVANCED
STUDIES

WORKING PAPERS

RSCAS 2020/30
Robert Schuman Centre for Advanced Studies
Global Governance Programme-393

Sources of Comparative Advantage
in Data-Related Services

Erik van der Marel

European University Institute

Robert Schuman Centre for Advanced Studies

Global Governance Programme

Sources of Comparative Advantage in Data-Related Services

Erik van der Marel

EUI Working Paper **RSCAS** 2020/30

Terms of access and reuse for this work are governed by the Creative Commons Attribution 4.0 (CC-BY 4.0) International license. If cited or quoted, reference should be made to the full name of the author(s), editor(s), the title, the working paper series and number, the year and the publisher.

ISSN 1028-3625

© Erik van der Marel, 2020

This work is licensed under a Creative Commons Attribution 4.0 (CC-BY 4.0) International license.
<https://creativecommons.org/licenses/by/4.0/>

Published in May 2020 by the European University Institute.
Badia Fiesolana, via dei Roccettini 9
I – 50014 San Domenico di Fiesole (FI)
Italy

Views expressed in this publication reflect the opinion of individual author(s) and not those of the European University Institute.

This publication is available in Open Access in Cadmus, the EUI Research Repository:
<https://cadmus.eui.eu>

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies, created in 1992 and currently directed by Professor Brigid Laffan, aims to develop inter-disciplinary and comparative research on the major issues facing the process of European integration, European societies and Europe's place in 21st century global politics.

The Centre is home to a large post-doctoral programme and hosts major research programmes, projects and data sets, in addition to a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration, the expanding membership of the European Union, developments in Europe's neighbourhood and the wider world.

For more information: <http://eui.eu/rscas>

The EUI and the RSCAS are not responsible for the opinion expressed by the author(s).

The Global Governance Programme

The Global Governance Programme is one of the flagship programmes of the Robert Schuman Centre. It is a community of outstanding professors and scholars that produces high quality research and engages with the world of practice through policy dialogue. Established and early-career scholars work on issues of global governance within and beyond academia, focusing on four broad and interdisciplinary areas: Global Economics, Europe in the World, Cultural Pluralism and Global Citizenship.

The Programme also aims to contribute to the fostering of present and future generations of policy and decision makers through its executive training programme: the Academy of Global Governance, where theory and 'real world' experience meet and where leading academics, top-level officials, heads of international organisations and senior executives discuss on topical issues relating to global governance.

For more information: <http://globalgovernanceprogramme.eui.eu>

The European University Institute and the Robert Schuman Centre are not responsible for the opinions expressed by the author(s).

Abstract

Does digital infrastructure increase trade benefits in services, and how large are these trade benefits? This paper uses the prevalence of countries' data centres and secure internet servers to investigate the impact of data-related infrastructures on services trade. Data centres and internet servers facilitate the production and trade of many service sectors reliant on software technologies. Therefore, digital infrastructures are complementary sources of comparative advantage in the more data-reliant sectors. Instrumental variable regressions underline this trade impact by using as instruments a country's exposure to natural hazards since 1900, such as earthquakes, floods and droughts. Overall, data infrastructures are an important and exogenous determinants of specialization patterns in services dependent on software.

Keywords

Trade in services; Data centres; Internet

JEL classification: F14; L86; H54

1. Introduction*

Data-related services traded cross-border have grown steadily since the early 2000s. The share of data-related services exports has more than doubled since 2004 and now represents a share exceeding 0.6 percent in global GDP. Moreover, non-digital services such as finance and retail have become more data-intensive and are increasingly traded over the internet, shifting away from their traditional supply of services through foreign affiliates (Andrenelli, 2018).

What determines the achievement of countries in producing and trading data-related services? Invariably, some countries are more successful in exporting these services than others and therefore enjoy comparative advantage in the data-related sectors. Technological innovations such as cloud computing and advanced information and telecommunication technology (ICT) have certainly contributed to the expansion of data-related services trade. However, one potential important factor in explaining the success of countries in exporting data-related services are data centers, internet exchange points (IXP) and internet servers. This paper therefore examines whether these data-related infrastructures form an exogenous and independent source of comparative advantage for countries exporting in data-related services.

Recent previous works have mainly concentrated on the regulatory environment in explaining the ability of countries to export data-related services. They therefore failed to attribute any importance to the role of specific digital infrastructures. Ferracane and van der Marel (2018) have shown that countries with lower levels of restrictiveness with respect to data, such as the prevention to apply data localization requirement, exhibit higher shares of trade in data-related services. Similarly, Goldfarb and Trefler (2018) explore key features of data in the form of Artificial Intelligence with respect to international trade and discuss the policy implications such as privacy, data localization and standards. Both works and other existing research therefore omit the likely impact of data-related infrastructures on data-related services trade, something this paper does.

Disentangling the importance of digital infrastructures to data-related services trade is important because descriptive data does not provide a straightforward answer. Figure 1 shows for instance that various lower middle-income countries have been extremely effective in exporting ICT-services. This in spite of the skewed distribution of data-related infrastructures around the globe, as shown in Figure 2. It is therefore unclear what role policy makers should attribute to the development of data centers, internet exchange points and secure internet servers. In great part these factors provide the backbone of many digital services exports *in addition* to an open policy regime in data. As a result, this paper makes a serious attempt in examining the question whether these types of digital infrastructures have any bearing on trading data-intensive services.

In doing so, this paper adds value to the previous literature in three specific ways. One, this paper collects data on the number of data servers that countries have built over the years, something that no other paper has done so far, to ascribe the importance of this specific type of infrastructure on services trade. Other types of data-related infrastructures which are also taken as a focal point in this paper are internet exchange points and secure internet servers. This point makes a distinctive difference compared to previous literature that examines a certain proxy of data flows to examine its contribution to trade (Manyika et al., 2016). Knowing that this source of data is bilateral-specific between countries, and because all data flows of within and between countries needs to pass through data-related infrastructures, this paper takes the latter approach to find out its relevance for data-related trade.

Second, and most importantly, we use instrumental variable strategy to find out how data-related infrastructures form a source of countries accomplishment in trading data-related services. Even though

* The author would like to thank Paola Conconi, Bernard Hoekman and Nicola Gagliardi for helpful comments

data flows and trade in services are hugely intertwined (WTO, 2019), instrumenting data flows is difficult. Instead, instrumenting digital infrastructures in our case has the convenient advantage of obtaining a valid instrument: Firms that construct data-related infrastructures such as data centres strongly take into account a country's exposure to the risk of natural disasters in their decision making. This information presents us with exogenous variation in order to assess the true impact of data infrastructure on data-related services trade. Our empirical results show that the selection of this instrument is a valid one.

Third, an older set of works have demonstrated that the internet is a strong predictor for cross-border trade in services (Freund and Weinhold, 2002; 2004). Although this literature reflects the line of research this paper takes, it is nonetheless markedly different. The distinctive difference is that over the years the digital infrastructure for trading services online has changed dramatically. Cloud computing, data storage technologies and other computer system resources were not available yet in the early 2000s or were otherwise deployed on a minimal scale. While the internet is necessary to transmit data, which are embedded in many digital services, it hardly defines the entire set of organizational facilities needed for a country to trade services digitally. Accompanying data-related infrastructures such as data centers to explain trade in services currently represents a novelty in the empirical academic literature.

The remainder of the paper is organized as follows. The next section discusses the previous literature with respect to data-related trade in services, or more broadly, digital services trade. Then, the third section sets out the simple empirical strategy in which data-intensive services trade are regressed on the three types of data-related infrastructures. The fourth section develops the instrumental variable strategy that is used in the empirical model and presents the results of the instrumental variable (IV) regressions. Finally, the last section concludes and puts the results in a wider context.

2. Previous Literature

The previous literature on data, trade in services, and data centers is scant. From a macro-perspective, the importance of data in international trade globally has been estimated by Manyika *et al.* (2016). The study states that cross-border data flows account for \$2.8 trillion of the total amount of increased world GDP over the last decade, thereby exerting a larger impact on growth than traditional goods trade. Interestingly, this work does not dedicate special attention to the inter-linkages that exist between data flows and trade in services. It takes the data as a separate type of flow that impacts the economy independent from services. That flow, however, has over the years increased dramatically. Besides the sheer flows of data that has increased exponentially since 2005, associated internet protocol (IP) traffic was estimated to have grown 64 times since 2005, whereas global internet bandwidth has quadrupled between 2010 and 2014 (Pepper *et al.*, 2014).

Earlier work from Freund and Weinhold (2002) points to the internet as a facilitating factor on trade in services. The authors show that an increase in internet penetration by 10 percent boosts growth of services imports by 1.1 percentage point and exports by 1.7 percentage point. These numbers illustrate that the internet effectively decrease transport costs for services as a subset of trade costs. Overall, the US ITC estimates that the internet reduces trade costs by about 26 percent on average (USITC, 2014), overcoming the historical hurdle of borders on trade. Lendle *et al.* (2016) find similar reductions in the border effect to be 64 percent for a comparable basket of goods that is likewise traded for a similar set of countries over the internet using eBay. In this case, the authors claim the reduction in trade costs comes from the way the internet cuts search costs given the effect rises with product differentiations and for countries speaking different languages.

However, positive trade costs developments for data and associated services also have the possibility to be reversed. Goldfarb and Trefler (2018) make clear that as Artificial Intelligence (AI) expands, data-related policy implications for services are likely to rise. For example, regulatory "behind-the-border" policies such as data localization and strict data privacy rules have the potential to limit the ability of

foreign firms to access data and to scale up their business models. This policy restriction may therefore inhibit countries' trade patterns and distort sources of comparative advantage. Recent work by Ferracane and van der Marel (2018) indeed shows that countries with higher levels of policy restrictions in data, such as data localization, exhibit lower levels of services trade. The authors show that this is particularly true for services that are data-intensive by means of software usage.

The involvement of data centres as critical digital infrastructure has been highlighted in various non-academic country and regional-specific case studies. Clipp et al. (2014) assess the impact of Facebook's settlement of a data centre in Sweden. The report claims that it has brought alongside an entirely new digital ecosystem worth more than the direct investment costs of the infrastructure itself. García Zaballos and Iglesias (2017) discuss how data centres suit the purpose of economic development for Latin American and Caribbean countries. They argue too that data centres are essential for the development of an ICT ecosystem in the region. The authors also consider a framework that portrays the optimal environment in which data centres are developed. One important factor that supports the decision to locate data centres in countries is the resilience to natural risk disaster – something this paper explores as an exogenous instrument.

On a larger scale this paper closely relates to a long line of research that describes more generally the impact of infrastructure on trade costs and comparative advantage. Donaldson (2018) estimates that physical infrastructure (in the form of railroads networks in India) shape comparative advantage in goods trade. Analogous to this paper, because countries differ in digital infrastructures, they also differ in productivity levels across services, which incites countries to trade in order to exploit comparative advantage. Stretching the concept of infrastructure as an endowment factor or rather country characteristic, this paper also relates to the works of Nunn (2007), Levchenko (2007), Costinot (2009) and eventually Chor (2011). These works develop an empirical approach also used in this paper. That is, they exploit differences in country factors such as quality of domestic institutions and per worker human capital as a source for comparative advantages in goods that are intensive in the factor. In our case, this “factor” data.

3. Empirical Strategy

This section sets out the empirical strategy. The empirical strategy uses an interaction term in which country-specific characteristics are multiplied with industry-level intensities. This interaction defines comparative advantage from where trade benefits arise, following the above-mentioned set of seminal works such as Nunn (2007), Levchenko (2007), Costinot (2009) and Chor (2011). In this paper, the interaction term is comprised of an indicator measuring a country's prevalence of data centres and other related data infrastructures, multiplied by the extent to which service sectors are intense in the use of software. The latter term is conveniently called sectoral data-intensities. This is how this paper's source of comparative advantage in data (or software) intensive services set up: by a country's level of digital infrastructure drawing on data centres and related types of data infrastructures which are discussed below.

Equation (1) measures formally how data infrastructure such as data centres form a country's source of comparative advantage which is associated with data-intensive trade. In particular, we regress the logarithm of cross-border exports of services (SX) in country c , for service sector j , on the interaction term that is comprised of a term called DC for country c and a term denoted by D/L for sector j . Hence, the empirical baseline model takes the following form:

$$\ln(SX)_{cj} = \Phi + \theta \ln(DC)_c * \ln(D/L)_j + \delta_c + \gamma_j + \varepsilon_{cj} \quad (1)$$

In equation (1) the terms $\ln(DC)_c$ and $\ln(D/L)_j$ form the multiplication of country characteristics and sector-intensities, which respectively is denoted by DC (data centres) and D/L (data-intensity). The term DC represents a vector in which we not only use number of data centres per country, but also equivalent

metrics of internet exchange points (IXP) and secure internet services (SIS) (see below). Then, the terms δ_c and γ_j refer to the fixed effects by exporter and sector, respectively. Sector fixed effects are applied at the 2-digit BPM6 level, which includes a total of 18 service sectors. Finally, ε_{cj} is the residual term. Regressions are estimated with robust standard error clustered by country-sector and are performed over the period 2016-2017 in a cross-sectional setting throughout. Hence, no time dimension is given to our identification strategy.

For services exports, $\ln(SX)_{cj}$, this work uses the WTO-UNCTAD-ITC annual trade in services dataset, which covers exports and imports of total commercial services. This is our preferred source which we also use for our descriptive analysis below. This dataset covers 222 entities which include countries and regional aggregations/economic groupings from 2005-2017 at the 2-digit level. The data is in line with the sixth edition of the IMF Balance of Payments and International Investment Position Manual (BPM6) as well as the 2010 edition of the Manual on Statistics of International Trade in Services (MSITS 2010). This entails that, compared to the BPM5 classification, major changes for the Balance of Payments (BOP) classification for services have been introduced with regards to financial intermediation services, insurance services, intellectual property and manufacturing and maintenance services.

3.1 Data Centres

In its simplest explanation, data centres are physical facilities that house data and enable digital network applications for external organizations. Often established in large buildings, or a set of buildings, data centres are dedicated spaces where computing capabilities and storage resources enable the delivery of shared applications for external organization. Data centres also store, manages and disseminates data for the applications. In effect, the data centre owns computer systems and associated components, such as telecommunication and storage systems, to organize an external organization's IT operations and equipment. Often, data centres accommodate data and the network application techniques that are critical for the external organization to the continue their daily digital operations. In simple terms, data centres are also commonly known as "the cloud", which are basically servers located somewhere that are connected via ICT networks.

As a result, reliability, network support and security assurance are important aspects to guarantee the operations of a data centre. To adhere to these qualities, the businesses literature identifies several criteria that firms consider in their decision to select a site for constructing a data centre. Three of the most important factors play a role in this decision-making process, namely environmental conditions, power supply and communications infrastructure. Regarding the first factor, it covers a site's climate and history of natural hazards – an item that we utilize for our instrumental variable strategy. Other secondary factors that play a role in firms' decisions to select a site are socio-economic conditions such as skilled workforce, the availability of construction services and an appropriate governance framework in terms of existing regulations.

Figure 3 shows that the global distribution of data centres is unequal, giving natural rise to cross-country differences of this data-related infrastructure. The prevalence of data centres in each country is measured per 1 Mln population, which is also used in our regressions. The map shows that countries such as Iceland, Latvia, Switzerland, Mauritius and Hong Kong have high densities of data centres. Most countries well-endowed with data centres are relatively smaller open-oriented economies. Two interesting countries with one of the highest data centre concentrations are Iceland and New Zealand; two countries that are geographically remote compared to many others. On the other side of the spectrum, countries showing lowest levels of data centre are Tanzania, Bangladesh, Peru, Afghanistan, but also China and Mexico. The last two countries that have somewhat higher income levels compared to others in the bottom ranking.

In addition to data centres, we also enter the number of internet servers (SIS) and internet exchange points (IXP) into the regressions. Internet servers are computer programs or devices that provides

functionality for other programs or devices, called "clients". Typical servers are database servers, file servers, mail servers, print servers, web servers, game servers, and application servers. Oftentimes, these servers are placed at data centres, but not always. Internet exchange points are physical infrastructure through which internet service providers (ISPs) and content delivery networks (CDNs) exchange internet traffic between their networks. See for the variable list for further information.

In the regression, all three data infrastructures are divided by 1Mln population to get a normalized measure. The reason for doing so is that firms take into account the size of the market when investing in each of the three types of physical infrastructures for data. Economies of scale plays an important role, but as outlined above, other factors too. Also, given that we are in a framework of specialization (i.e. comparative advantage), we need to normalize this country measure. At a later stage we also perform robustness checks by dividing the number of the three data infrastructures by employment and land area (sq km). We take the log for each measure after dividing so that in fact for data centres we use $\ln(DC/P)_c$, for secure internet servers we get $\ln(SIS/P)_c$, and for internet exchange points the term becomes $\ln(IXP/P)_c$.

3.2 Data-intensities

Data-intensities are measured using information on software usage. Specifically, this paper takes the 2011 Census ICT Survey from the US, which reports data at detailed 4-digit NAICS sector level. This data is survey-based and records how much each industry and service sector spend in Mln USD on ICT technology in terms of hardware equipment and computer software.

The survey records two types software expenditure: capitalised and non-capitalised. We select capitalised expenditure given that this proxy is closer to the concept of intensities with respect to labor and capital as a factor of production used in the academic literature (e.g. Romalis, 2004; Chor, 2011). Non-capitalised expenditure instead relates more to the input support of firms which enters in the production function as intermediate inputs. Capitalised expenditure is comprised of longer-term investments made in computer software. It excludes purchases and payroll for developing software as well as software licensing and service, and maintenance agreements for software. The year 2010 is selected so that this information doesn't run the risk of being endogenous to the trade data. Capitalized software expenditure is divided over labour, for which we also use data for the year 2010.

This proxy for "data-intensity" is not ideal. However, there is currently no good data on how much data is used by each sector. Several sources such as Cisco and Telegeography guesstimate the extent to which data exist within countries, but only for a handful of observations, not sectors. Having said that, what is clear, and what is also intuitively appealing, is that the use of data within and across borders is performed using software technologies. Firms need software to use the internet in its simplest form and employ advanced software to transmit large sets of data that connect to data centres. In addition, more technology advanced transmissions of data over the internet are done with the help of cloud computing technologies that data centres provide and which in themselves are extremely software intense. As such, software capital is in our view the first-best available proxy to date.

Intensities are computed at 4-digit NAICS level and then concorded into 2-digit BPM6 level given that the trade data recorded in this classification system. Because no concordance table exists between NAICS and BPM6, a self-constructed matrix is used. Numbers are aggregated at 2-digit BPM6 level by taking the simple average. Note that one sector forms a mismatch between the two classification tables, which is Intellectual property / Royalties and license fees. This category is not reported in the US Census nor in the BLS database. Nonetheless, this sector is important as it covers, among other items, patents, trademarks and copyrights – activities which are data-intensive and for which the trade data records high

levels of services exports. Therefore, we have developed our own concordance table to include this sector. Details of this procedure can be found in Annex 1.¹

Table 1 ranks the service sectors by software intensity. Unsurprisingly, telecommunications, computer services and information services exhibit highest software usage. These sectors employ a high amount of data by means of software usage compared to labour. Information services cover activities that are closely related to the work of a data centre, such as data processing services, but also includes web search. Both financial and insurance services are high data-intensive sectors too. The two sectors are more broadly considered as very digital intensive given that over the years internet technologies have massively changed the financial services industry.² On the other side of the spectrum, the least software-intensive sectors are construction and travel services. The middle-range is made up of a mix of modern and traditional sectors such as R&D and transport services.

Figure 4 illustrates the thrust of this paper's hypothesis. It shows that the prevalence of data centres in countries is strongly associated with trade in data-intensive services. The figure takes the sector of information services as an example, but equally strong correlations hold for other data-intensive sectors. Countries with a higher level of data centres per 1 Mln people demonstrate greater shares of information services exports in GDP, such as Switzerland, UK or Singapore. Conversely, Guatemala, Kazakhstan and Peru are countries with low shares of information services exports as part of their economies but also having low densities of data centres.

4. Results Baseline Regressions

Tables 2 and 3 report the empirical results of the baseline regressions. Besides the number of data centres (DC), we also provide results from the two additional variables in the DC vector as given in equation (2), namely for internet exchange points (IXP) and secure internet servers (SIS).

In Table 2, the first column shows the coefficient result using data centres as the variable of interest. Columns 2 and 3 report the results for using IXPs and secure internet servers. In all three occasions the coefficients are positive and strongly significant, indicating that countries which have a higher prevalence of data centres do indeed exhibit higher volumes of exports in more data-intensive service sectors. Similarly, countries which are more supplied with IXPs and secure internet servers tend to export more in data-intensive services. These findings remain robust for data centres and secure internet servers when including GDP per capita (in PPP) which measures a country's level of development, as reported in Table 2. The empirical literature makes clear the inclusion of this control variable reinforces the case that the variable of interest identifies an independent channel through which trade patterns are determined.

Regarding the comparable importance between the two remaining significant variables, yielding standardized beta coefficients provides a gauge of their relative impacts. In doing so, results of the beta coefficient of secure internet servers is more than three times as large ($\beta_{SIS} = 0.399$) compared to the beta coefficient found for data centres ($\beta_{DC} = 0.129$).³ However, one must keep in mind that on average

¹ The concordance table between 4-digit NAICS and 2-digit BPM6 can be obtained upon request. Admittedly, the inclusion of intellectual property / royalties and license fees as a service is a BOP decision and some debate exists whether this is truly a service. In addition, for some countries, this may also reflect tax and transfer pricing as drivers of observable trade in this sector. However, since this sector is included in all publicly available data sources recording trade in services, we prefer to include it. Nonetheless, in our regression we have also dropped this sector entirely as additional (unreported) robustness checks. Results do not alter in any way apart from slight coefficient size changes. Results are available and can be obtained upon request.

² Another non-ICT sector that is shown to be very software-intensive is the retail sector. However, neither the US Census nor the BPM6 classification shows a separate entry for retail or wholesale distribution services, which is the reason why this sector is omitted in our analysis of intensities and is not covered in our regression analysis.

³ Results are not reported but can be obtained upon request.

countries have more than 240.000 secure internet servers on their territory against 40 data centres. Therefore, even though a one standard deviation increase in secure internet servers is more than 3 times as effective, in terms of actual units this outcome differs: a 10 percent increase in data centres, equivalent to 4 units, results in an increase of about 0.5 percent in exports of data-related services. The corresponding percentage increase in secure internet servers, which is 24.000, results in an increase of 0.42 percent of data-intensive services exports.

4.1 Endogeneity

One great concern with our empirical specification is that a higher prevalence of digital infrastructures is observed in countries that trade more data-related services. In other words, reverse causality is an obvious probability. In order to ensure that the significance of data centres and secure internet servers are not a function of services trade, this paper uses the frequency rate of natural hazards of countries as an instrument. A country's natural hazard rate is caused by a geophysical, meteorological, hydrological and climatological onset events. They represent strong external forces that isolate any potential influence resulting from existing trade patterns.⁴

The use of natural onsets is an intuitively appealing instrument for in particular data centres. Firms deciding on a site to construct data centres generally consider many factors. One of the top criteria for the site selection are the environmental conditions. Especially long-term weather conditions prevail in the decision making for constructing a data centre, such as wind, snow and ice storms, as well as other natural hazards which are composed of sudden onsets such as seismic events, floods, tornados, hurricanes and volcanos. These weather conditions leave little room for any control and are unlikely to be influenced by trade. Therefore, the number of onset occurrences, or more broadly natural hazards, as an instrument for data centres is a strong exogenous factor with respect to a country's export patterns in data-related services.

Especially Long-term meteorological conditions are an important determinant in the decision-making process. Firms avoid places with excessive wind patterns as well as snow and ice storms. More generally, data centres are often placed by considering the optimum temperature of a location and country's relative humidity. Taking into account these onset indicators is essential, because the digital communication networks and hardware IT equipment are designed for operating only within a certain range of temperatures and humidity. Also, cooling techniques that data centres extensively make use of depend on environmental conditions as they affect energy demands for air conditioning. Precipitation on the other hand appears less of a problem as firms can deal with such weather conditions by protecting the data centre.

In addition, other natural hazards also play a large role in deciding for a location. These are composed of more sudden onsets, namely hydrological natural hazards which are caused by the occurrence, movement, and the distribution of surface and subsurface freshwater and saltwater. Examples are floods that are of a great concern for data centre developers. Interestingly, firms also like to stay away from locations near oceans because of high-sulphates and natural corrosives (salt) that can damage the data centre. Finally, geophysical natural hazards matter too for data centres, which can manifest in the form of earthquakes and volcanic activity. Notably volcanos are of concern for data centre constructors. Volcanic eruptions create a large amount of ashes that in combination with strong wind patterns can cause long-term damage to data centres.

Of note, a country's natural hazard rate caused by a geophysical, meteorological, hydrological and climatological onset event also matter for internet servers as they are often, but not always, placed at the

⁴ There are also biological and extra-terrestrial natural hazards. For instance, epidemic or insect infestation are examples of the former, whereas space weather is an example of the latter. Both categories are omitted here.

premise of the data centre. As such, our instrument also deals the risk of endogeneity with respect to internet servers.

4.2 Instrument

We take data on natural hazards from the EM-DAT database. This data source is also known as the International Disaster Database from the Centre for Research on the Epidemiology of Disasters (CRED). The database records two types of hazards, namely natural hazards and technology hazards. This paper takes natural hazards for creating the instrument.⁵

Natural hazards consist of geophysical, meteorological, hydrological, climatological, as well as biological and extra-terrestrial onset events. This paper takes the first four types of natural hazards as a focal point for creating the instrument. It therefore omits the last two types of natural hazards.⁶ The database only report disasters related to these hazards, which means that an onset event only enters the database if one of the following three criteria is recorded: (1) 10 or more people deaths, (2) 100 or more people affected, injured or homeless, or (3) a declaration exits by the country of a state of emergency and/or an appeal for international assistance.⁷ The database records information by country, region and precise location, and also documents temporal information.

Geophysical hazards are earthquakes (including tsunamis), mass movements and volcanic activity, which are all onset events that originate from solid earth. Meteorological hazards are storms, extreme temperatures and fogs; they are generally onsets caused by short-lived extreme weather and atmospheric conditions that can last for days. Hydrological hazards are onsets caused by the occurrence, movement, and distribution of surface and sub-surface of freshwater and saltwater which appears as floods, landslides and wave actions. Finally, climatological hazards are onsets caused by atmospheric processes stemming from seasons or climate variability, such as drought and wildfire. This latter category also covers slow-moving onsets.

To create a variable, the four categories of hazards are taken together. Specifically, the occurrence rate is computed by taking the sum of each of the four categories of hazards, as well as for each category separately, since 1900-2015. This number for each country is then divided by its 2015 population so as a per Mln rate is obtained. The reason for taking a long historical record is because a lengthier time horizon makes any estimation using this instrument even less sensitive to current services trade influences – even though reverse causality is unlikely to happen in the first place.⁸ The reason for dividing over population is that the goal of this paper is to find out specialization patterns, in line with the aforementioned comparative advantage literature. Besides, firms naturally take market size into account when deciding on an investment.⁹

Figure 4 shows the correlation between the instrument of onset occurrences and data centres per 1Mln population in logs. A relatively tight relationship appears. The figure indicates that countries with higher per unit levels of data centres also exhibit greater per unit levels of onset occurrences. Note that

⁵ Technological hazards are comprised of industrial accidents, such as gas leaks and oil spills, as well as transport accidents and other miscellaneous accidents such as collapses and explosions.

⁶ Biological natural hazards are epidemics, insect infections and animal accidents. Extra-terrestrial natural hazards are airbursts and more generally space weather such as geomagnetic storms or shock waves.

⁷ The database reports that in some cases secondary criteria are also taken into account when figures are missing, such as the significance of the disaster over a time span or whether significant damage takes place.

⁸ One could argue, for instance, that past services trade performance is correlated with overall higher levels of economic activity which brings along onset event in society. Given that this paper takes an extended time horizon of natural hazards since 1900, it is very unlikely that services trade today impacts this variable.

⁹ Robustness checks have also been performed using total employment and land area as de denominator. Results are largely consistent and can be obtained upon request.

the correlation is positive, which at first may look counter-intuitive. Yet the unit levels using population neutralizes the size of each country that otherwise would be picked up if nominal values of onset occurrences were used: bigger countries also experience more hazards. By construction, then, smaller sized economies such as Iceland, Hong-Kong, Bulgaria and Mauritius naturally show greater levels of data centre endowments together with hazard frequencies.

4.3 Results IV Regressions

The results of the IV regressions are reported in Tables 4 and 5 for data centres and secure internet servers respectively. These were the two significant variables in the baseline regressions. In column 1 of each table the instrument variable of occurrence rate (OR) is computed using all four types of natural hazards occurrences together, as explained above, and which is titled “Total”. In the second column, only the category of meteorological hazards are taken up to compute the occurrence rate and are denoted with “Meteo” in the table.

In Table 4, the results in the first two columns have a positive coefficient sign, in line with the initial correlation of Figure 5. The coefficient size for both regressors more or less double in each column compared to the results in Table 2. Both columns also report an F-statistic that is much higher than 10, indicating the instrument is strong. In addition, the p-value of the Kleibergen–Paap rk LM-statistic shows that the instrument is also relevant as we can reject the null hypothesis of under-identification. Note that the coefficients for the first stage regressions are also positive and significant. However, the endogeneity test rejects the OR Total instrument in column 1 whereas the OR Meteo instrument only pass the endogeneity test at a marginal level ($p \leq .10$).

Further, in columns 3 and 4 of Table 4 the regressions are repeated but now with GDP per capita. Including this control variable is problematic by default: a higher level of development is responsive to a country’s higher levels of services trade (Francois and Hoekman, 2010; Hoekman and Mattoo, 2008). Therefore, in the IV regressions, we treat this variable as endogenous too. Doing so gives in both columns a positive and significant coefficient outcome for the variable of interest, although with half of the coefficient size consistent with the results found in the baseline regressions. In both cases the p-values of the Sargan–Hansen J-statistic tell us that the joint null hypothesis is not rejected. The two instruments are therefore both valid and uncorrelated with the error term.

The results reported in Table 5 are largely similar as found in Table 4. In all columns the coefficient results for secure internet servers are highly significant, experience a reduced coefficient size when the GDP term is entered, and are measured with a strong instrument. This stays true when instrumenting secure internet services with both the total and meteorological onset occurrence rate.

However, the treatment of the per capital GDP variable remains somehow unsatisfactory. Even though the addition of this control variable makes the case stronger for our coefficient results not to be already picked up by levels of development, ideally this covariant should be treated as exogenous. Finding a way to make GDP per capita uncorrelated with services trade and the instrument uncorrelated with the error is therefore essential. This would also make the IV estimator consistent. One solution is to take past economic performance which is unlikely to be affected by current levels of services trade. Moreover, given that we take an extreme time horizon for onset occurrences, historical data on per capita GDP levels are also unlikely to influence this instrument. As a result, the year 2007 is chosen which is a time that lies well ahead of our cross-section.

Table 5 reports the results when treating GDP per capita levels for the year 2007 as exogenous. The first two columns show that when instrumenting for data centres, the regressions generates significant outcomes. Similarly, when instrumenting for secure data servers the last two columns also show significant outcomes, although the coefficient result for OR Meteo is statistically weak. Note that in all cases the GDP variable is negative yet insignificant. One potential explanation is that on average less developed countries have grown faster since 2007 which has resulted in greater levels of services trade

as opposed to richer countries, which have seen a slower growth rate in GDP. In all cases, the instrument is strong (F-statistic > 10) although the endogeneity test is only rejected in columns 2 and 4, signifying that instrumentation is necessary.

On the basis of these results in column 2 and 4, a 10 percent increase in data centres now results in an expansion of exports in data-related services of about 1.6 percent, which is equivalent to Singapore's share in world exports in services. Likewise, an equal percentage increase in secure internet servers raises data-related exports with 2.1 percent, comparable to the relative size of total services export of a country like Canada.

4.4 Other Robustness Checks

This section performs various robustness checks. First, we take employment and land area as the denominator instead of population. Second, we check for alternative years for the per capital GDP variable. Third, we check whether our results are not driven by the sheer levels of trade. Fourth, we use an index of cross-border data flow restrictiveness as an additional control variable.

The first robustness check is to see whether results hold when replacing total population for total employment or land area. So far population is used to normalize our measure of data centres (DC) and secure internet servers (SIS) prevalence because firms take into account market size when investing in the construction of a data centre. Using employment instead would be in line with our (D/L) variable of software-intensity and be somewhat closer to an interpretation based on factor proportions following Romalis (2004). One should know that strictly speaking digital infrastructure such as data centres do not form any endowment factor in the production function. It therefore cannot be seen as an Heckscher-Ohlin determinant of comparative advantage (Chor, 2011). Still, using employment is informative to see whether results are sensitive to different numerators.

Tables 7 and 8 report the results for data centres and secure internet servers respectively when using total employment. Employment is sourced from the Penn World Tables for the year 2017. In Table 7, three out of four columns give a significant result of the regressions for data centres. The result is particularly strong when using OR Meteo as instrument. Column 4 shows that the results remain significant when entering the GDP variable. The results for secure internet servers are also significant albeit weaker. Here, no significant results are obtained when using the GDP term as a control variable. Note that if we take total land area (sq. km) instead of employment, results are largely similar but statistically weaker as shown in Tables 9 and 10.

The second robustness check is to make sure that the year for the per capita GDP data is not arbitrary chosen. It may be that using the year 2007 for the control variable gives a significant result by chance. It may also be that 2007 is too close to our cross-section year of 2017 and so the GDP term might still be sensitive to any reverse causality. This might in particular be true when the effect of economic development involves a substantial time lag for countries to experience higher levels of services trade. Therefore, we take the years 2001, 2004, in addition to 2010 as an extra year to see whether results hold. Table 11 and 12 shows that for all years, and for both data centres and secure internet servers, results are consistently significant, although for earlier years the coefficients are statistically weaker.

The following robustness check is to see whether there's truly a shift away from existing exports in favour of data-related services. It may be that sheer trade as a response variable does not entirely capture the relative specialization pattern of a country. Countries may have higher trade levels in data-related services, but that wouldn't necessarily imply a relative change in their export basket vis-à-vis already exported services. Even though the previous literature, e.g. Chor (2011), precisely tests comparative advantage by the interaction term as defined in the baseline specification, we nonetheless test this issue. Specifically, we compute the Balassa index of Revealed Comparative Advantage (RCA) based on our services trade data and substitute this indicator in equation (1).

The idea behind this indicator is that it summarizes the concept of Ricardian comparative advantages in one measure. However, given that this measure has received various critique as it is asymmetric, i.e. unbounded for those sectors with an RCA higher than 1, but a zero as a lower bound for sectors with a comparative disadvantage, we normalize the index following standard practise as proposed by Laurson (2015) with $(RCA - 1) / (RCA + 1)$. The results are reported in Table 13 for data centres and Table 14 for secure internet servers. The outcomes for data centres are significant throughout all specifications except in column 3. Importantly, the coefficient result is significant when using OR Meteo as instrument in addition to the GDP variable for the year 2007 as a control. The results for secure internet services remain insignificant when entering the GDP term.

The fourth and final robustness check is to control for policy restrictions in digital services. Cross-border services trade as picked up in this paper appears to be susceptible to restrictions that prohibit or inhibit the movement and domestic use of data. In particular, Ferracane and van der Marel (2018) show that restrictions to the free flow of data across borders is negatively associated with trade in data-intensive services. Given that this paper takes data centres and secure internet servers as a focal point, of particular interest is their sub-index of cross-border data restrictions. Data centres as well as internet services, which are mostly integrated in data centres, operate by the existence of cross-border data flows and therefore have an impact on data-related services trade.

The restrictiveness indicator is sourced from ECIPE where it forms part of the larger Digital Trade Restrictiveness Index (DTRI). Only restrictions to the cross-border movement of data is selected, and so restrictions regarding the domestic use of data is excluded. Ferracane and van der Marel (2018) show that the latter category of restrictions has little impact on services trade. Cross-border data flow restrictions include data localization rules, local storage requirement and restrictive conditional flow regimes. To ensure that the restrictiveness index is not endogenous to services trade, the same approach as with the GDP variable above is applied: a year well-ahead of the cross-section period is taken, namely the year 2010. We do this for both the restrictiveness index and the GDP variable.

Results for data centres are reported in Table 15, whereas results for secure internet servers are reported in Table 16. In both tables the coefficient results in columns 1 and 2 are significant. It means that the two instruments stay robust when including policy restrictions on data. Interestingly, the coefficient results on the data policy restrictiveness index also comes out as statistically significant, which appears to be stronger when using OR Meteo as instrument. It implies that lower levels of cross-border data flow restrictions form a complementary factor for enhancing trade in data-related services alongside developing digital infrastructures. Note that this corresponding result is stronger for secure internet servers than for data centres.

A final interesting result in column 3 and 4 of both tables is that the results show a coefficient results on data restrictions that remains negatively significant when entering the per capita GDP variable.

5. Conclusion and Policy Implication

This paper shows that data-developed countries are more successful exporters in service sectors more reliant on software. The paper illustrates that countries with higher rates of data centres and secure internet servers are in a better position to specialize in services which are software intense. This paper uses historical occurrence rates of natural hazards such as meteorological onsets as an instrument to show that this impact is robust and exogenous. Data infrastructures are sensitive to extreme onset occurrences which firms like to avoid. In sum, countries with a higher prevalence of data centres and internet servers, which are likely to experience more favourable weather conditions, are in a better position to export in digital services.

Differing productivity levels across service sectors create incentives for countries to trade in order to stimulate comparative advantage. One way of doing so is to reduce restrictive policies related to the cross-border movement and domestic usage of data. Many emerging and developing countries would

benefit from such a reform approach as generally these countries have shown greatest increase in data-related restrictions. However, that cannot be the only factor for policy makers to consider. Indeed, the empirical analysis this paper develops makes a strong case that the digital infrastructure is a robust enabling factor that complements data policies.

Yet, increasing digital infrastructure capacities has its limits. Particularly for developing countries building data centres is an expensive undertaking. Also, small open economies may not always want to prioritize the construction of data centres given their limited land area. In that context, the regional and global network can play a supportive role in developing data-related activities to reinforce comparative advantage. Countries that do not have enough capacity to construct data centres will rely on other countries for their digital infrastructures on which data can be stored. Yet, for that to happen there is a need to have a policy regime in which data can flow freely across borders. Regional or global frameworks to discuss data policies can therefore help to achieve the benefits of data-related services trade for these types of countries.

Obviously, then, digital policies and digital infrastructure go hand-in-hand to create greatest productivity benefits. In that regard, it is encouraging to see that robustness checks in this paper confirm that both aspects can contribute significantly in reaping gains arising from services trade: one through policy reforms and a second one through spending on digital infrastructures. However, this outcome does not provide policy makers with the complete story, as it leaves enough room for further research. For instance, it is far from clear how the two factors interact. What is the optimal sequencing model for governments to undertake? What can poorer countries do to increase disaster risk resilience given that many of them suffer from continued high rates of natural hazards?

Bibliography

- Andrenelli, A., et al. (2018), "Multinational production and trade in services", OECD Trade Policy Papers, No. 212, OECD Publishing, Paris.
- Chor, D. (2011) "Unpacking Sources of Comparative Advantage: A Quantitative Approach", *Journal of International Economics*, Vol. 82. No. 2, pages 152-167.
- Clipp, C., H. Larby, F. Lind, A. Lundmark, M. Morin (2014) "Digital Infrastructure and Economic Development: An Impact Assessment of Facebook's Data Centre in Northern Sweden", Boston Consultancy Group.
- Costinot, A. (2009) "On the Origins of Comparative Advantage", *Journal of International Economics*, Vol. 77, Issue 2, pages 255-264.
- Donaldson, D. (2018) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure", *American Economic Review*, Vol. 108, No. 4-5, pages 899-934.
- Ferracane, M.F. and E. van der Marel (2018) "Do Data Flows Restrictions Inhibit Trade in Services?", ECIPE DTE Working Paper Series No. 2, Brussels: ECIPE.
- Ferracane, M.F., H. Lee-Makiyama and E. van der Marel (2018b) "The Digital Trade Restrictiveness Index", ECIPE, Brussels.
- Ferracane, M.F., J. Kren and E. van der Marel (2018a) "Do Data Policy Restrictions Impact the Productivity Performance of Firms and Industries?", ECIPE DTE Working Paper Series No. 1, Brussels: ECIPE.
- Francois, J. and B. Hoekman (2010) "Services Trade and Policy", *Journal of Economic Literature*, Vol. 48, No. 3, pages 642-692.
- Freund, C. and D. Weinhold (2002) "The Internet and International Trade in Services" *American Economic Review*, Vol. 92, No. 2, pages 236-40.
- Freund, C. and D. Weinhold (2002) "The Internet and International Trade in Services", *American Economic Review*, Vol. 92. No. 2, pages 236-240.
- Freund, C. and D. Weinhold (2004) "The effect of the Internet on international trade", *Journal of International Economics*, Vol. 62, No. 1, pages 171-189.
- García Zaballos, A. and E. Iglesias Rodriguez (2017) "Data Centers and Broadband for Sustainable Economic and Social Development: Evidence from Latin America and the Caribbean" Inter-American Development Bank.
- Goldfarb, A. and D. Trefler (2018) "AI and International Trade", NBER Working Paper Series No. 24254, National Bureau of Economic Research, Cambridge MA: NBER.
- Hoekman, B. and A. Mattoo (2008) "Services Trade and Growth", World Bank Policy Research Working Paper Series 4461, Washington DC, The World Bank.
- Laursen, K. (2015) "Revealed Comparative Advantage and the Alternatives as Measures of International Specialization", *Eurasian Business Review*, Vol. 5, No. 1, pages 99 – 115.
- Lendle, A., M. Olarreaga, S. Schropp and P-L. Vézina (2016) "There Goes Gravity: eBay and the Death of Distance", *Economic Journal*, Vol. 126, Issue, 591, pages 406-441.
- Levchenko, A. (2007) "Institutional Quality and International Trade", *Review of Economic Studies*, Vol. 74, Issue 3, pages 791-819.
- Manyika, J., S. Lund, J. Bughin, J. Woetzel, K. and D. Dhingra (2016) "Digital Globalization: The New Era of Global Flows", McKinsey Global Institute, Washington DC: McKinsey and Company.

- Nunn, N. (2007) “Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade”, *The Quarterly Journal of Economics*, Vol. 122, Issue 2, pages 569-600.
- Robert Pepper, R., J. Garrity and C. LaSalle (2016) “Cross-Border Data Flows, Digital Innovation, and Economic Growth”, *The Global Information Technology Report 2016*, Chapter 1.2, WEF, Davos.
- Romalis, J. (2004) “Factor Proportions and the Structure of Commodity Trade”, *American Economic Review*, Vol. 94, No. 1, pages 67-97.
- USITC (2014) “Digital Trade in the U.S. and Global Economies, Part 2”, Publication 4485, Investigation No. 332-540, August. Washington, DC: US International Trade Commission.
- WTO (2019) “World Trade Report 2019: The Future of Services Trade”, World Trade Organization, Geneva.

Variable definitions

Exports (SX): Exports of services traded cross-border. Trade flows are reported for each 2-digit BOP sector for each country. Year: 2016. Source: WTO-ITC-OECD.

Data-intensity (D/L): Expenditures on capitalized computer equipment and software over labor use (2010). Variables are calculated for each 4-digit NAICS sector and re-concorded for each 2-digit BMP6 sector using self-constructed concordance table. Year: 2010. Source: US Census Information and Communication Technology Survey (ICTS) and US Bureau of Labor Statistics.

Data centers (DC/P): Number of data centers that house computer systems and associated components, and which centralizes an organization's IT operations and equipment, and which also store, manage and disseminate data. Variable is divided by population in order to get a per unit basis using data from the World Development Indicators. Year: 2017/2018. Source: DataCenterMap.com.

Secure internet servers (SIS/P): The number of distinct, publicly trusted Transport Layer Security (TLS) / and its now-deprecated predecessor Secure Sockets Layer (SSL) certificates found in the Netcraft Secure Server Survey. This variable is divided by total population in order to get a per unit basis using data from the World Development Indicators. Year: 2018. Source: World Development Indicators.

Internet Exchange Points (IXP/P): Physical infrastructure through which Internet service providers (ISPs) and content delivery networks (CDNs) exchange Internet traffic between their networks. This variable is divided by total population in order to get a per unit basis using data from the World Development Indicators. Year: 2017/2018. Source: DataCenterMap.com.

Occurrence Rate (OR/P) Total: Occurrence Rate of geophysical, meteorological, hydrological, climatological natural hazards divided over 2015 population. Year: 1990-2015. Source: EM-DAT Natural Disasters Database.

Occurrence Rate (OR/P) Meteo: Occurrence Rate of meteorological natural hazards divided over 2015 population. Year: 1990-2015. Source: EM-DAT Natural Disasters Database.

GDP per capita (GDPpc PPP): Gross Domestic Product per capita based on purchasing power parity (PPP), current international USD Year: 2007. Source: World Development Indicators.

Population (P): Population, total. Year: 2015. Source: World Development Indicators.

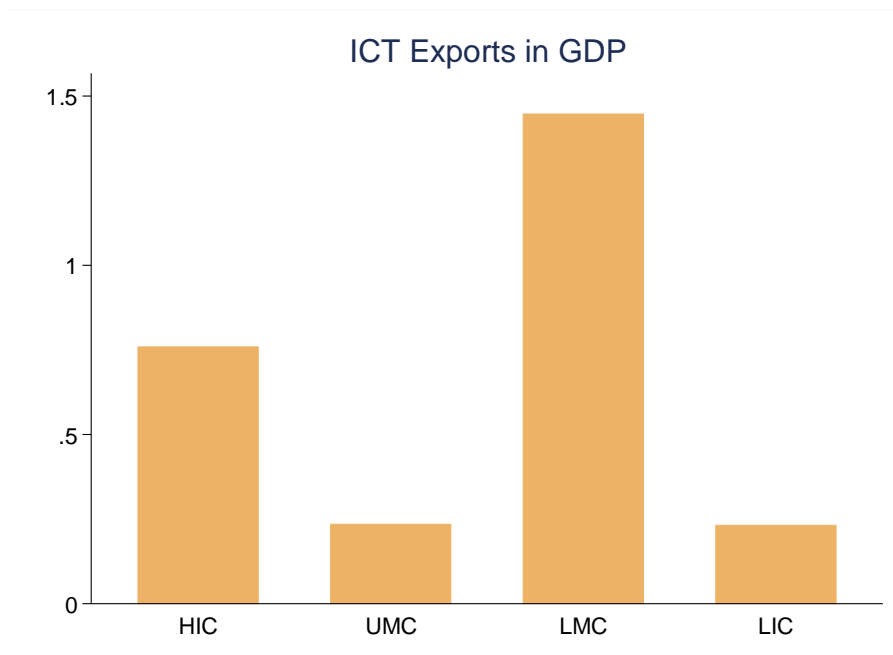
Employment (L): Number of persons engaged, total. Year: 2015. Source: Penn World Table 9.1.

Land sq. km (T): Land area, total, excluding area under inland water bodies, national claims to continental shelf, and exclusive economic zones. Year: 2015. Source: World Development Indicators.

DTRI CB: Digital Trade Restrictiveness Index Cross Border, which covers restrictive measures related to the cross-border data flows only such as data localization requirement, local storage requirement and conditional flow regimes. Year: 2010. Source: ECIPE.

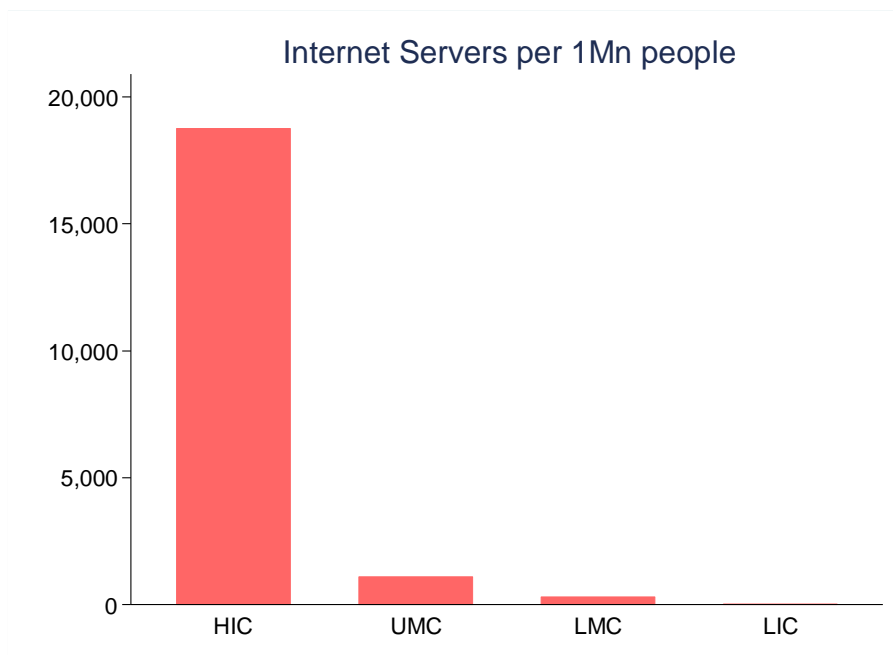
Tables and Figures

Figure 1: ICT Services Exports in Gross Domestic Product (GDP), 2018



Source: World Development Indicators. HIC: High-income Group; UMC: Upper Middle-income Group; LMC: Lower Middle-income Group; LIC: Low-income Group.

Figure 2: Secure Internet Servers per 1 Mln Population, 2018.



Source: World Development Indicators. HIC: High-income Group; UMC: Upper Middle-income Group; LMC: Lower Middle-income Group; LIC: Low-income Group.

Figure 3: World Map of Data Centres per 1Mln Population



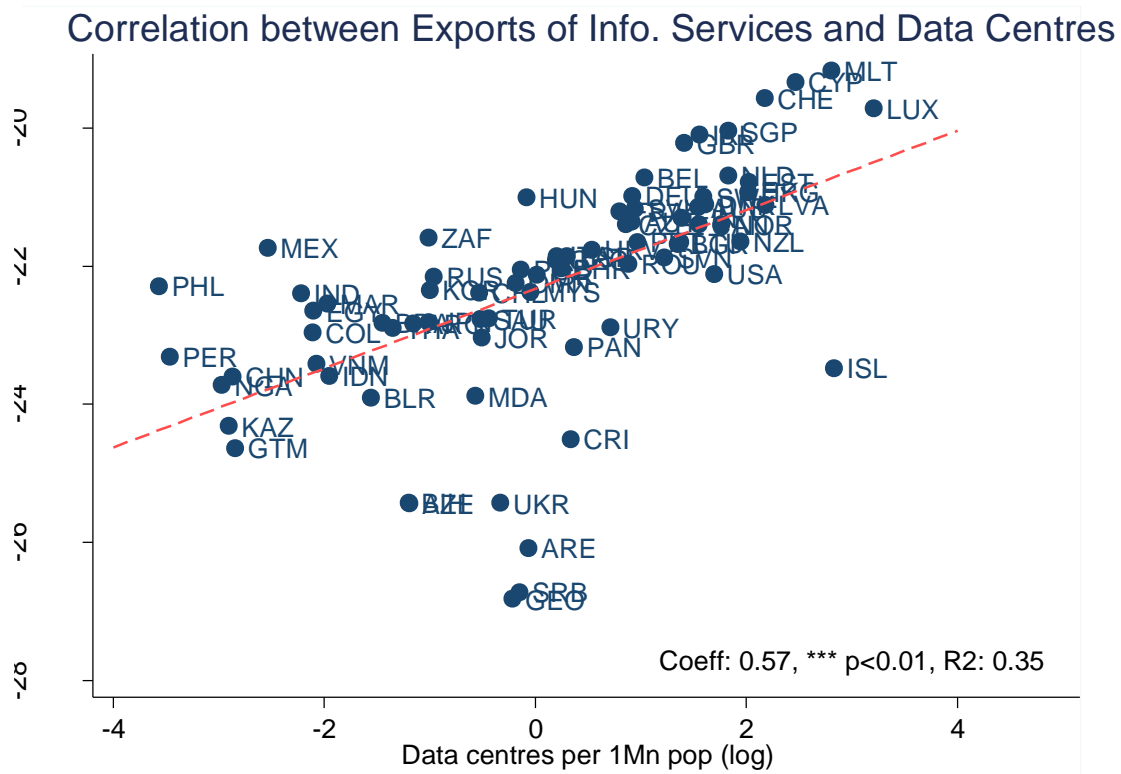
Source: Data Center Map

Table 1: Ranking of Data-intensive (D/L) services by 2-digit BPM6 sector (18 sectors)

2-digit BPM6 code	Sector Description	ln(D/L)
SI1	Telecommunications services	5.917
SI2	Computer services	4.556
SI3	Information services	3.359
SG	Financial services	3.040
SF	Insurance and pension services	2.306
SH	Charges for the use of intellectual property n.i.e.	1.615
SC3	Other modes of transport	1.413
SC1	Sea transport	1.169
SK1	Audio-visual and related services	0.946
SC2	Air transport	0.894
SJ1	Research and development services	0.696
SJ2	Professional and management consulting services	0.547
SC4	Postal and courier services	0.369
SK2	Other personal, cultural, and recreational services	0.309
SJ3	Technical, trade-related, and other business services	0.304
SB	Maintenance and repair services n.i.e.	0.136
SE	Construction	0.062
SD	Travel	0.050

Source: US Census & US BLS; see Variable Definitions

Figure 4: Correlation between Information Services Exports in GDP and Data Centres per 1Mln Population



Source: WTO-UNCTAD-ITC and World Bank World Development Indicators, Data Center Map

Table 2: Baseline regression results

	(1)	(2)	(3)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(DC/P)_c$	0.067*** (0.000)		
$\ln(D/L)_j * \ln(IXP/P)_c$		0.043*** (0.006)	
$\ln(D/L)_j * \ln(SIS/P)_c$			0.050*** (0.000)
FE Exporter	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes
Observations	1519	1071	1902
R2A	0.867	0.889	0.871
R2W	0.019	0.009	0.020
RMSE	1.263	1.047	1.344

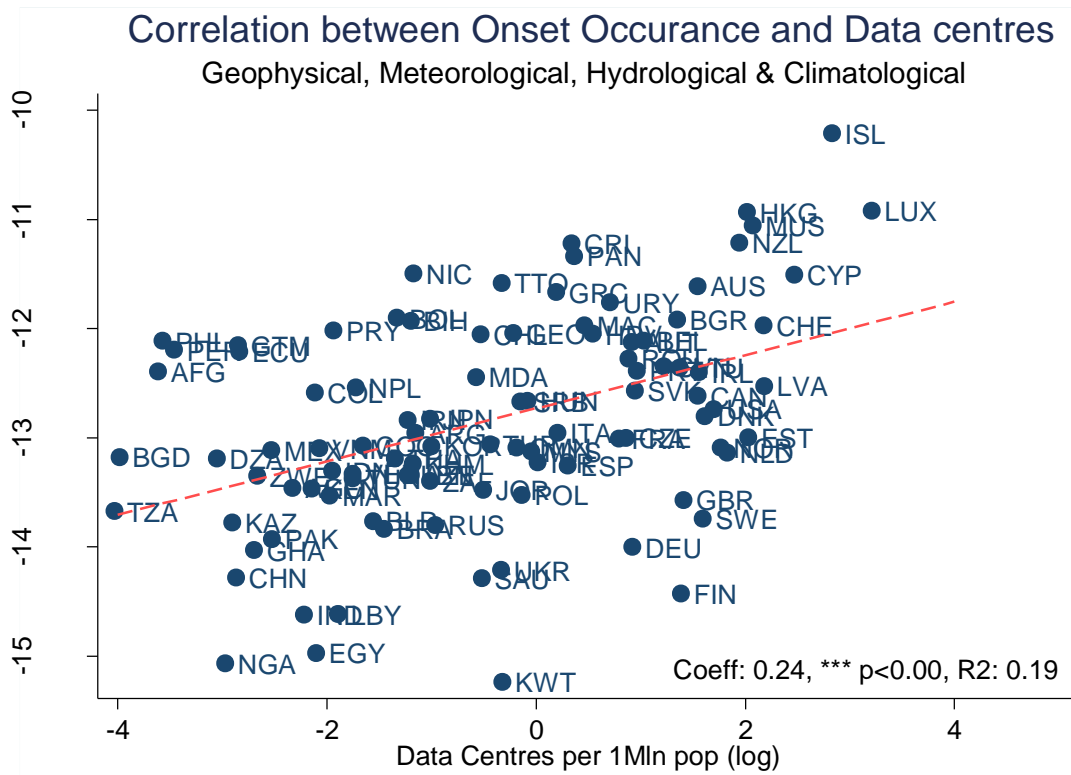
Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. IXP denote Internet Exchange Points. SIS denotes Secure Internet Servers.

Table 3: Baseline regression results with control variable

	(1) ln(SX) _{cj}	(2) ln(SX) _{cj}	(3) ln(SX) _{cj}
ln(D/L) _j * ln(DC/P) _c	0.048** (0.047)		
ln(D/L) _j * ln(IXP/P) _c		-0.021 (0.328)	
ln(D/L) _j * ln(SIS/P) _c			0.042*** (0.003)
ln(D/L) _j * ln(GDPpc PPP) _c	0.049 (0.396)	0.194*** (0.001)	0.004 (0.926)
FE Exporter	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes
Observations	1426	1071	1778
R2A	0.877	0.891	0.880
R2W	0.018	0.033	0.016
RMSE	1.218	1.035	1.302

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services exports ln(SX) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term ln(D/L) denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. IXP denote Internet Exchange Points. SIS denotes Secure Internet Servers.

Figure 5: Correlation between Onset Occurrences and Data Centres per 1Mln People



Source: EM-DAT database, World Bank World Development Indicators, Data Center Map

Table 4: IV regression results for Data Centres (DC)

	(1) ln(SX) _{cj}	(2) ln(SX) _{cj}	(3) ln(SX) _{cj}	(4) ln(SX) _{cj}
ln(D/L) _j * ln(DC/P) _c	0.107*** (0.001)	0.117*** (0.000)	0.075*** (0.000)	0.074*** (0.001)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1412	1345	1350	1283
R2	0.881	0.879	0.888	0.887
RMSE	1.187	1.153	1.161	1.123

First stage regressions

	OR Total	OR Meteo	OR Total	OR Meteo
ln(D/L) _j * ln(OR/P) _c	0.797*** (0.000)	0.778*** (0.000)	0.463*** (0.000)	0.416*** (0.000)
ln(D/L) _j * ln(GDPpc PPP) _c			1.461*** (0.000)	1.265*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	273.206	567.688	745.723	511.142
p-value Chi-sq Test	0.001	0.000	0.001	0.003
p-value End. Test	0.332	0.082	0.413	0.482
p-value J-statistic	-	-	0.618	0.755

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services exports ln(SX) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term ln(D/L) denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 5: IV regression results for Secure Internet Servers (SIS)

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(SIS/P)_c$	0.117*** (0.008)	0.106*** (0.004)	0.043*** (0.001)	0.044*** (0.001)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1801	1650	1713	1562
R2	0.015	0.004	0.016	0.013
RMSE	1.281	1.242	1.234	1.199

First stage regressions

	OR Total	OR Meteo	OR Total	OR Meteo
$\ln(D/L)_j * \ln(OR/P)_c$	0.452*** (0.000)	0.488*** (0.000)	0.342*** (0.000)	0.184*** (0.000)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c$			2.210*** (0.000)	2.132*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	42.080	85.652	613.633	575.505
p-value Chi-sq Test	0.006	0.003	0.003	0.005
p-value End. Test	0.109	0.099	0.969	0.729
p-value J-statistic	-	-	0.557	0.675

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 6: IV regression results using GDP per capita PPP 2007

	(1) ln(SX) _{cj}	(2) ln(SX) _{cj}	(3) ln(SX) _{cj}	(4) ln(SX) _{cj}
ln(D/L) _j * ln(DC/P) _c	0.121** (0.021)	0.162*** (0.004)		
ln(D/L) _j * ln(SIS/P) _c			0.154** (0.044)	0.218* (0.076)
ln(D/L) _j * ln(GDPpc PPP) _c '07	-0.041 (0.611)	-0.114 (0.182)	-0.201 (0.182)	-0.321 (0.179)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1410	1343	1773	1622
R2	0.881	0.878	0.029	0.094
RMSE	1.189	1.158	1.283	1.287

First stage regressions

	OR Total	OR Meteo	OR Total	OR Meteo
ln(D/L) _j * ln(OR/P) _c	0.504*** (0.000)	0.441*** (0.000)	0.272*** (0.000)	0.160*** (0.000)
ln(D/L) _j * ln(GDPpc PPP) _c '07	1.325*** (0.000)	1.148*** (0.000)	1.943*** (0.000)	1.898*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	205.593	186.964	22.280	12.469
p-value Chi-sq Test	0.019	0.003	0.025	0.044
p-value End. Test	0.228	0.070	0.107	0.089

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services exports ln(SX) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term ln(D/L) denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 7: IV regression results for Data Centres (DC) using employment (L)

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(DC/L)_c$	0.056** (0.043)	0.085*** (0.001)	0.045 (0.242)	0.096** (0.037)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			0.057 (0.375)	-0.028 (0.702)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1405	1340	1388	1323
R2	0.885	0.884	0.885	0.884
RMSE	1.158	1.120	1.161	1.125

First stage regressions

	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/L)_c$	0.716*** (0.000)	0.753*** (0.000)	0.526*** (0.000)	0.461*** (0.000)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			1.283*** (0.000)	1.094*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	147.967	357.635	175.056	146.821
p-value Chi-sq Test	0.048	0.001	0.246	0.038
p-value End. Test	0.712	0.354	0.999	0.321

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. L denotes employment (Labour). DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 8: IV regression results for Secure Internet Servers (SIS) using employment (L)

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(SIS/L)_c$	0.077* (0.054)	0.072** (0.022)	0.084 (0.164)	0.125 (0.219)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			-0.076 (0.524)	-0.152 (0.437)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1705	1560	1696	1551
R2	0.011	0.009	0.010	0.013
RMSE	1.238	1.201	1.230	1.204
<i>First stage regressions</i>				
	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/L)$	0.457*** (0.000)	0.563*** (0.000)	0.301*** (0.000)	0.182*** (0.001)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			1.902*** (0.000)	1.840*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.001
F-statistic	30.169	76.424	18.294	10.144
p-value Chi-sq Test	0.054	0.023	0.152	0.203
p-value End. Test	0.437	0.350	0.501	0.360

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. L denotes Employment (Labour). SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 9: IV regression results for Data Centres (DC) using land (T)

	(1)	(2)	(3)	(4)
	ln(SX)	ln(SX)	ln(SX)	ln(SX)
$\ln(D/L)_j * \ln(DC/T)_c$	0.034 (0.138)	0.043** (0.041)	0.025 (0.315)	0.030 (0.222)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			0.103** (0.014)	0.085* (0.056)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1412	1345	1410	1343
R2	0.880	0.878	0.881	0.879
RMSE	1.192	1.157	1.187	1.153

First stage regressions

	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/T)_c$	0.893*** (0.000)	0.823*** (0.000)	0.797*** (0.000)	0.705*** (0.000)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			1.227*** (0.000)	0.931*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	750.259	1075.753	1016.001	859.128
p-value Chi-sq Test	0.139	0.041	0.316	0.222
p-value End. Test	0.255	0.972	0.907	0.382

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. T denotes Land (Terra). DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 10: IV regression results for Secure Internet Servers (SIS) using land (T)

	(1)	(2)	(3)	(4)
	ln(SX)	ln(SX)	ln(SX)	ln(SX)
$\ln(D/L)_j * \ln(SIS/T)_c$	0.036*	0.039*	0.032	0.031
	(0.083)	(0.071)	(0.179)	(0.269)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			0.040	0.047
			(0.405)	(0.408)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	0.017	0.014	0.016	0.014
R2	1.256	1.225	1.249	1.216
RMSE	1.256	1.225	1.249	1.216

First stage regressions

	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/T)$	0.737***	0.660***	0.639***	0.498***
	(0.000)	(0.000)	(0.000)	(0.000)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			1.889***	1.733***
			(0.000)	(0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	220.295	286.352	249.337	196.854
p-value Chi-sq Test	0.083	0.070	0.178	0.268
p-value End. Test	0.774	0.968	0.774	0.501

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. T denotes Land (Terra). SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 11: IV regression results for Data Centres (DC) using GDP per capita PPP 2001-2010

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
	GDP 2001	GDP 2004	GDP 2007	GDP 2010
$\ln(D/L)_j * \ln(DC/P)_c$	0.117** (0.025)	0.119** (0.021)	0.121** (0.021)	0.121** (0.018)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c$	-0.022 (0.764)	-0.032 (0.671)	-0.041 (0.611)	-0.044 (0.595)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1404	1410	1410	1410
R2	0.880	0.881	0.881	0.881
RMSE	1.189	1.189	1.189	1.189

First stage regressions

	Total	Total	Total	Total
$\ln(D/L)_j * \ln(OR/P)_c$	0.507*** (0.000)	0.509*** (0.000)	0.504*** (0.000)	0.516*** (0.000)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c$	1.191*** (0.000)	1.261*** (0.000)	1.325*** (0.000)	1.367*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	198.702	205.855	205.593	232.541
p-value Chi-sq Test	0.024	0.020	0.019	0.017
p-value End. Test	0.205	0.222	0.228	0.236

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 12: IV regression results for Secure Internet Servers (SIS) using GDP per capita PPP 2001-2010

	(1) ln(SX) _{cj} GDP 2001	(2) ln(SX) _{cj} GDP 2004	(3) ln(SX) _{cj} GDP 2007	(4) ln(SX) _{cj} GDP 2010
ln(D/L) _j * ln(SIS/P) _c	0.163* (0.058)	0.160* (0.052)	0.154** (0.044)	0.145** (0.035)
ln(D/L) _j * ln(GDPpc PPP) _c	-0.193 (0.215)	-0.201 (0.197)	-0.201 (0.182)	-0.192 (0.167)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1767	1773	1773	1773
R2	0.047	0.037	0.029	0.020
RMSE	1.294	1.288	1.283	1.277

First stage regressions

	Total	Total	Total	Total
ln(D/L) _j * ln(OR/P) _c	0.244*** (0.000)	0.255*** (0.000)	0.272*** (0.000)	0.300*** (0.000)
ln(D/L) _j * ln(GDPpc PPP) _c	1.781*** (0.000)	1.871*** (0.000)	1.943*** (0.000)	2.014*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	16.686	18.783	22.280	28.726
p-value Chi-sq Test	0.033	0.030	0.025	0.020
p-value End. Test	0.104	0.107	0.107	0.110

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable the log of services exports ln(SX) using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term ln(D/L) denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 13: IV regressions results for Data Centres (DC) using RCA

	(1)	(2)	(3)	(4)
	RCA _{cj}	RCA _{cj}	RCA _{cj}	RCA _{cj}
$\ln(D/L)_j * \ln(DC/P)_c$	0.022** (0.012)	0.031*** (0.000)	0.023 (0.102)	0.038*** (0.007)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			0.003 (0.909)	-0.017 (0.478)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1390	1327	1388	1325
R2	0.021	0.026	0.022	0.024
RMSE	0.349	0.342	0.349	0.343
<i>First stage regressions</i>				
	Total	Meteo	Total	Meteo
$\ln(D/L) * OR$	0.792*** (0.000)	0.787*** (0.000)	0.514*** (0.000)	0.450*** (0.000)
$\ln(D/L) * \ln(GDPpc\ PPP)_c '07$			1.354*** (0.000)	1.183*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	256.969	523.194	204.583	192.174
p-value Chi-sq Test	0.013	0.000	0.103	0.008
p-value End. Test	0.968	0.272	0.767	0.213

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable is the normalized RCA indicator using data from the WTO-UNCTAD-ITC BPM6 database and has been cleaned for extreme outliers. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 14: IV regression results for Secure Internet Servers (SIS) using RCA

	(1)	(2)	(3)	(4)
	RCA _{cj}	RCA _{cj}	RCA _{cj}	RCA _{cj}
$\ln(D/L)_j * \ln(SIS/P)_c$	0.022 (0.113)	0.026** (0.013)	0.026 (0.228)	0.044 (0.167)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '07$			-0.021 (0.633)	-0.050 (0.433)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	1769	1623	1741	1595
R2	0.016	0.018	0.010	0.011
RMSE	0.373	0.367	0.372	0.369
<i>First stage regressions</i>				
	Total	Meteo	Total	Meteo
$\ln(D/L) * OR$	0.423*** (0.000)	0.467*** (0.000)	0.274*** (0.000)	0.168*** (0.000)
$\ln(D/L) * \ln(GDPpc\ PPP)_c '07$			1.934*** (0.000)	1.903*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	35.207	74.120	21.953	13.091
p-value Chi-sq Test	0.113	0.013	0.214	0.150
p-value End. Test	0.608	0.372	0.496	0.295

Note: * p<0.10; ** p<0.05; *** p<0.01. The dependent variable is the normalized RCA indicator using data from the WTO-UNCTAD-ITC BPM6 database and has been cleaned for extreme outliers. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database.

Table 15: IV regression results for Data Centres (DC) and DTRI CB

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(DC/P)_c$	0.039** (0.043)	0.040** (0.029)	0.004 (0.945)	-0.002 (0.971)
$\ln(D/L)_j * \text{DTRI CB}_c '10$	-0.333 (0.198)	-0.406* (0.063)	-0.410* (0.079)	-0.479** (0.026)
$\ln(D/L)_j * \ln(\text{GDPpc PPP})_c '10$			0.108 (0.379)	0.118 (0.288)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	980	954	980	954
R2	0.881	0.875	0.882	0.876
RMSE	0.875	0.859	0.870	0.854
<i>First stage regressions</i>				
	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/P)_c$	1.008*** (0.000)	0.959*** (0.000)	0.430*** (0.000)	0.449*** (0.000)
$\ln(D/L)_j * \text{DTRI CB}_c '10$	7.173*** (0.000)	3.975*** (0.000)	1.828*** (0.003)	0.928 (0.120)
$\ln(D/L)_j * \ln(\text{GDPpc PPP})_c '10$			1.740*** (0.000)	1.487*** (0.000)
p-value LM Statistic	0.000	0.000	0.000	0.000
F-statistic	345.521	464.390	84.879	68.643
p-value Chi-sq Test	0.046	0.029	0.945	0.971
p-value End. Test	0.528	0.295	0.536	0.463

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. DC denotes Data Centres. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database. CB denotes Cross-Border and covers cross-border data flow restrictions using ECIPE's DTRI.

Table 16: IV regression results for Secure Internet Servers (SIS) and DTRI CB

	(1)	(2)	(3)	(4)
	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$	$\ln(SX)_{cj}$
$\ln(D/L)_j * \ln(SIS/P)_c$	0.051** (0.050)	0.046** (0.030)	0.086 (0.728)	-0.004 (0.971)
$\ln(D/L)_j * DTRI\ CB_c '10$	-0.578* (0.054)	-0.502** (0.032)	-0.585* (0.075)	-0.474* (0.064)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '10$			-0.077 (0.880)	0.123 (0.620)
FE Exporter	Yes	Yes	Yes	Yes
FE Sector	Yes	Yes	Yes	Yes
Observations	991	954	991	954
R2	0.008	0.005	-0.003	0.016
RMSE	0.891	0.859	0.896	0.854
<i>First stage regressions</i>				
	Total	Meteo	Total	Meteo
$\ln(D/L)_j * \ln(OR/P)_c$	0.754*** (0.000)	0.850*** (0.000)	0.095* (0.083)	0.194*** (0.001)
$\ln(D/L)_j * DTRI\ CB_c '10$	7.641*** (0.000)	5.634*** (0.000)	1.142* (0.083)	1.722*** (0.006)
$\ln(D/L)_j * \ln(GDPpc\ PPP)_c '10$			1.974*** (0.000)	1.908*** (0.000)
p-value LM Statistic	0.000	0.000	0.078	0.000
F-statistic	97.042	209.509	3.002	11.864
p-value Chi-sq Test	0.052	0.029	0.722	0.971
p-value End. Test	0.499	0.249	0.768	0.880

Note: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. The dependent variable the log of services exports $\ln(SX)$ using data from the WTO-UNCTAD-ITC BPM6 database. Robust standard errors are clustered at the country-industry level. Fixed effects for sector are applied at 2-digit BPM level. The term $\ln(D/L)$ denotes the log of data intensity over labour using data from US Census on capitalized software and US BLS. P denotes Population. SIS denotes Secure Internet Servers. OR denotes Occurrence Rate of natural hazards (Total or Meteo) using the EM-DATA database. CB denotes Cross-Border and covers cross-border data flow restrictions using ECIPE's DTRI.

Annex 1: The category of Royalties and licenses & Intellectual property

The category of Royalties and license and Intellectual property are two different names that refer to the same variable which are found in WTO-UNCTAD-ITC and OECD-WTO (BaTIS) trade in services databases. In the WTO-ITC-UNCTAD database, to which we refer as BPM6, this category is called Intellectual property whereas in the BaTIS database this category is denoted as Royalties and licenses.

Unfortunately, no direct connection between the NAICS 2007 classification and the sectors Royalties and licenses nor Intellectual property can be made from where we have computed our data intensities, i.e. (D/L). Equally unfortunate is that no concordance table exists between NAICS and BPM6 and NAICS and EBOPS more generally. Therefore, we have constructed our own concordance tables and build them up from an extremely detailed 6-digit level. This is not too difficult when mapping each 6-digit NAICS code into a 2-digit BPM6 or EBOPS code. However, since no clear 6-digit NAICS code can be directly linked to the services category of Royalties and license or Intellectual property, we have extended our concordance scheme to include this sector. We have done so in an indirect way through other concordance systems. The result of this concordance process can be seen in Table A1.1 below.

The way to do so is not clear-cut and some assumptions need to be made. For starters, the WTO-UNCTAD-ITC trade in services database designates Intellectual Property as chapter “SH” following the 6th edition of the Balance of Payments (BPM6) while the OECD-WTO BaTIS denotes this category as S266 following EBOPS 2002. As said, both overlap and are therefore indicated as “SH / S266” in Table A1.1. To eventually arrive at the NAICS 2007 code, two sequential sources are needed. First, the Annex III of the MSITS 2002 EBOPS classification provides a concordance table between EBOPS and CPC 1.0, which is used as a first step. Four sectors are classified under 266 Royalties and license fees, namely Patents, Trademarks, Copyrights and Other non-financial intangible assets. With the help of the United Nations correspondence tables website (<https://unstats.un.org/unsd/classifications>), a concordance can be made between CPC 1.0 and finally NAICS 2007 through five successive steps as outlined in Table A1.1.

Many different NAICS 2007 codes fall into one of the four original CPC 1.0 codes and therefore not all of them are equally relevant for Royalties and licenses or Intellectual property services. For that reason, we are not taking all 6-digit NAICS 2007 which eventually trace back to the two BPM6 and EBOP 2002 sectors as given in Table A1.1. The reason is that not all NAICS 2007 sectors are fully covered by the two intangible sectors. We only identify those which are not partially covered. These sectors are given in bold in column “NAICS 2002 / 07” of Table A.11 and are not given an * under the column “P” (which stands for partial). The information on whether an item is covered partially or not also comes from the United Nations correspondence tables. To come up with 2-digit BPM6 and EBOPS 2002 sector intensities, we take the unweighted average of each data intensity of these designated non-partial NAICS 2007 sectors, which should give us eventually a good approximation of the level of data used in the two sectors of Royalties and license and Intellectual property.

As one can see, a mix of service sectors fall under the two sectors, namely R&D services, some financial services, as well as cultural services such as motion pictures and sound recording. Also trust funds are fully covered under this category of Royalties and license / Intellectual property. Of note, the NAICS sector 515120 is not included under EBOPS, but is covered under BPM6 following their respective manuals.

Table A1.1: Concordance for Intellectual property & Royalties and license fees

BPM6 / EBOPS	CPC 1.0	CPC 1.0 description	CPC Prov.	CPC Ver.1.1	ISIC Rev.3	ISIC Rev.3.1	NAICS 2002 / 07	P	NAICS 2002 / 07 description
SH / S266	51210	Patents	89210	81110 / 81120 / 81130 / 81140 / 81150 / 81190 / 81300	7310	7310	541710		R&D in agriculture, electronics, environmental, biology, botany, biotechnology, computers, chemistry, etc.
SH / S266	51220	Trademarks	89220	73340	6599	6599	523110		Acting as a principle in the underwriting or dealing of securities
SH / S266							523130	*	Acting as a principal in the buying or selling of commodity futures or spot market contracts
SH / S266							523910	*	Venture capital companies and own account investors
SH / S266							523991	*	Administration of trusts and private estates
SH / S266							533110		Patent holders, franchise issuers, and holders of similar intangible assets receiving royalties or licensing fees
SH / S266							551111		Bank holding companies
SH / S266							551112	*	Other financial holding companies (except bank holding companies)
SH / S266							813211		Foundations or charitable trusts raising funds and providing grants to applicants with specific qualifications or specific institutions
SH / S266	51230	Copyrights	89230	73310	7220	7221	511210		Software publishing
SH / S266				73320	9211 / 9249	9211	512110	*	Motion picture production
SH / S266							512120		Motion picture and video distribution to theatres, television and cable networks and stations, and exhibitors
SH / S266							512191		Post production services for motion pictures and video, such as editing, dubbing, subtitling, titling, animation, special effects, closed captioning, and similar
SH / S266							512199	*	Motion picture film laboratories and booking agencies
SH / S266							512240		Sound recording studios
SH / S266							512290	*	Recording books on tape (except publishers)
SH / S266					9213	9213	512110	*	Production of television shows, commercials, etc., not done in broadcast facilities
SH / S266							512290	*	Production of radio programs (except live) not done in broadcast facilities
SH / S266							515111	*	Radio networks, including transmission of purchased or self-produced content

Erik van der Marel

SH / S266					515112	*	Radio stations, including transmission of purchased or self-produced content
SH / S266					515120	*	Television broadcasting, including transmission of purchased or self-produced content
SH / S266					515210	*	Cable networks, including transmission of purchased or self-produced content
SH / S266			9214	9214	561599	*	Ticket agencies, theatrical
SH / S266					711110	*	Theatre companies, comedy troupes, opera companies, live theatrical production
SH / S266					711120		Dance companies, groups, and ballets
SH / S266					711130		Musicians and musical groups
SH / S266					711190	*	Other performing arts companies, such as magic shows, ice skating shows, and similar
SH / S266					711310	*	Arts events organizers with facilities, arts facilities operators
SH / S266					711320	*	Arts events organizers without facilities
SH / S266					711510	*	Artists, writers, speakers, journalists, actors, cartoonists, dancers, producers, art restorers (all independent).
SH / S266	51290	Other non-financial intangible assets	89290	73390	6599	523110	Acting as a principle in the underwriting or dealing of securities
SH / S266					523130	*	Acting as a principal in the buying or selling of commodity futures or spot market contracts
SH / S266					523910	*	Venture capital companies and own account investors
SH / S266					523991	*	Administration of trusts and private estates
SH / S266					533110		Patent holders, franchise issuers, and holders of similar intangible assets receiving royalties or licensing fees
SH / S266					551111		Bank holding companies
SH / S266					551112	*	Other financial holding companies (except bank holding companies)
SH / S266					813211		Foundations or charitable trusts raising funds and providing grants to applicants with specific qualifications or specific institutions

Source: United Nations, BPM6 and EBOPS 2002.

Author contacts:

Erik van der Marel

Senior Economist at ECIPE,

Lecturer at the Université Libre de Bruxelles (ULB) and

Consultant Economist at the World Bank

Avenue des Arts 40

B-1050 Brussels

Email: erik.vandermarel@ecipe.org



With the support of the
Erasmus+ Programme
of the European Union

The European Commission supports the EUI through the European Union budget. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.